(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2008 (12.09.2008)

PCT

(10) International Publication Number
WO 2008/107112 A2

(51) International Patent Classification:
G06K 9/00 (2006.01) *G06K 9/62* (2006.01)

(21) International Application Number:
PCT/EP2008/001578

(22) International Filing Date:
28 February 2008 (28.02.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/892,881 5 March 2007 (05.03.2007) US

(71) Applicant *(for all designated States except US)*: **FOTONATION VISION LIMITED** [IE/IE]; Galway Business Park, Dangan, Galway (IE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BLONK, Leendert** [NL/IE]; Corofin, Tuam, County Galway (IE). **CORCORAN, Peter** [IE/IE]; Cregg, Claregalway, Galway (IE). **GANGEA, Mihnea** [RO/RO]; Sos. Nicolae Titulescu Nr. 10, Bloc 20, Scara A et 9, Ap. 44, Sector 1, Bucuresti (RO).

(74) Agents: **BOYCE, Conor** et al.; F.R. Kelly & Co, 27 Clyde Road, Ballsbridge, Dublin 4 (IE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL,

[Continued on next page]

(54) Title: ILLUMINATION DETECTION USING CLASSIFIER CHAINS

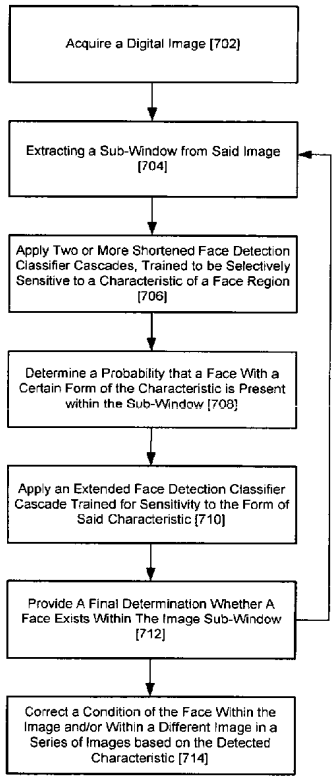


Figure 9A

(57) Abstract: A face illumination normalization method includes acquiring a digital image including a face that appears to be illuminated unevenly. One or more uneven illumination classifier programs are applied to the face data to determine the presence of the face within the digital image and/or the uneven illumination condition of the face. The uneven illumination condition may be corrected to thereby generate a corrected face image appearing to have more uniform illumination, for example, to enhance face recognition.

WO 2008/107112 A2

NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

Illumination Detection Using Classifier Chains

FIELD OF THE INVENTION

The invention relates to face detection and recognition, particularly under uneven illumination conditions

DESCRIPTION OF THE RELATED ART

Viola-Jones proposes a classifier chain consisting of a series of sequential feature detectors. The classifier chain rejects image patterns that do not represent faces and accepts image patterns that do represent faces.

A problem in face recognition processes arises when faces that are unevenly illuminated are distributed in a large area of face space making correct classification difficult. Faces with similar illumination tend to be clustered together and correct clustering of images of the same person is difficult. It is desired to be able to detect faces with uneven illumination within images, or where another difficult characteristic of a face exists such as a face having a non-frontal pose. It is also desired to have a method to normalize illumination on faces, for example, for use in face recognition and/or other face-based applications.

SUMMARY OF THE INVENTION

A face illumination normalization method is provided. A digital image is acquired including data corresponding to a face that appears to be illuminated unevenly. One or more uneven illumination classifier programs are applied to the face data, and the face date is identified as corresponding to a face. An uneven illumination condition is also determined for the face as a result of the applying of the one or more uneven illumination classifier programs. The uneven illumination condition of the face is corrected based on the determining to thereby generate a corrected face image appearing to have more uniform illumination. The method also includes electronically storing, transmitting, applying a face recognition program to, editing, or displaying the corrected face image, or combinations thereof.

A face recognition program may be applied to the corrected face image. The detecting of the face and the determining of the uneven illumination condition of the face may be performed simultaneously. A set of feature detector programs are applied to reject non-face data from being identified as face data.

A front illumination classifier program may be also applied to the face data. An illumination condition may be determined based on acceptance of the face data by one of the classifier programs. The digital image may be one of multiple images in a series that include the face, and the correcting may be applied to a different image in the series than the digital image within which the illuminating condition is determined.

The uneven illumination classifier programs may include a top illumination classifier, a bottom illumination classifier, and one or both of right and left illumination classifiers. A front illumination classifier program may be applied to the face data. Two or more full classifier sets may be applied after determining that no single illumination condition applies and that the face data is not rejected as a face.

A face detection method is also provided. The face detection method includes acquiring a digital image and extracting a sub-window from the image. Two or more shortened face detection classifier cascades are applied that are trained to be selectively sensitive to a characteristic of a face region. A probability is determined that a face with a certain form of the characteristic is present within the sub-window. An extended face detection classifier cascade is applied that is trained for sensitivity to the certain form of the characteristic. A final determination is provided that a face exists within the image sub-window. The method is repeated one or more times for one or more further sub-windows from the image and/or one or more further characteristics.

The characteristic or characteristics may include a directional illumination of the face region, an in-plane rotation of the face region, a 3D pose variation of the face region. a degree of smile, a degree of eye-blinking, a degree of eye-winking, a degree of mouth opening, facial blurring, eye-defect, facial shadowing, facial occlusion, facial color, or facial shape, or combinations thereof.

The characteristic may include a directional illumination, and an uneven illumination condition may be determined by applying one or more uneven illumination classifier cascades. A front illumination classifier cascade may also be applied. An illumination condition of a face may be determined within a sub-window based on acceptance by one of the classifier cascades. The digital image may be one of multiple images in a series that include the face, and an uneven illumination condition of the face may be corrected within a different image in the series than the digital image within which the illuminating condition is

determined. An uneven illumination classifier cascade may include a top illumination classifier, a bottom illumination classifier, and one or both of right and left illumination classifiers.

A further face detection method is provided that includes acquiring a digital image and extracting a sub-window from said image. Two or more shortened face detection classifier cascades may be applied that are trained to be selectively sensitive to directional facial illumination. A probability may be determined that a face having a certain form of directional facial illumination is present within the sub-window. An extended face detection classifier cascade may be applied that is trained for sensitivity to the certain form of directional face illumination. A final determination is provided that a face exists within the image sub-window. The method may be repeated one or more times for one or more further sub-windows from the image and/or one or more further directional facial illuminations.

The digital image may be one of multiple images in a series that include the face, and an uneven illumination condition of the face may be corrected within a different image in the series than the digital image within which the illuminating condition is determined.

The uneven illumination classifier cascades may include a top illumination classifier, a bottom illumination classifier, and one or both of right and left illumination classifiers. A front illumination classifier cascade may also be applied. An illumination condition of a face may be determined within a sub-window based on acceptance by one of the classifier cascades.

A digital image acquisition device is also provided including an optoelectonic system for acquiring a digital image, and a digital memory having stored therein processor-readable code for programming the processor to perform any of the face detection illumination normalization methods described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram illustrating the principle components of an image processing apparatus according to a preferred embodiment of the present invention;

Figure 2 is a flow diagram illustrating the operation of the image processing apparatus of Figure 1; and

Figures 3A-3D shows examples of images processed by the apparatus of the preferred embodiment.

Figure 4 is a block diagram of an image processing system in accordance with certain embodiments.

Figure 5 illustrates a main image sorting/retrieval workflow in accordance with certain embodiments.

Figure 6A illustrates an exemplary data storage structure for an image collection data set.

Figures 6B and 6D illustrate aspects of an image classifier where the feature vectors for individual patterns can be determined relative to an "averaged" pattern (mean face) and where feature vectors for individual patterns are determined in absolute terms (colour correlogram), respectively.

Figures 6C and 6E illustrate the calculation of respective sets of similarity measure distances from a selected classifier pattern to all other classifier patterns within images of the Image Collection.

Figure 6F illustrates how multiple classifiers can be normalized and their similarity measures combined to provide a single, similarity measure;

Figure 7 is a block diagram of an in-camera image processing system according to certain embodiments.

Figure 8 illustrates a face illumination normalization method in accordance with certain embodiments.

Figure 9A-9B illustrate face detection methods in accordance with certain embodiments.

Figures 10A-10B illustrate a further method in accordance with certain embodiments.

DETAILED DESCRIPTION

Figure 1 illustrates subsystems of a face detection and tracking system according to certain embodiments. The solid lines indicate the flow of image data; the dashed line indicates control inputs or information outputs (e.g. location(s) of detected faces) from a module. In this example an image processing apparatus can be a digital still camera (DSC), a video

camera, a cell phone equipped with an image capturing mechanism or a hand held computer equipped with an internal or external camera.

A digital image is acquired in raw format from an image sensor (CCD or CMOS) [105] and an image subsampler [112] generates a smaller copy of the main image. A digital camera may contain dedicated hardware subsystems to perform image subsampling, for example, to provide preview images to a camera display and/or camera processing components. The subsampled image may be provided in bitmap format (RGB or YCC). In the meantime, the normal image acquisition chain performs post-processing on the raw image [110] which may include some luminance and color balancing. In certain digital imaging systems, subsampling may occur after post-processing, or after certain post-processing filters are applied, but before the entire post-processing filter chain is completed.

The subsampled image is next passed to an integral image generator [115] which creates an integral image from the subsampled image. This integral image is next passed to a fixed size face detector [120]. The face detector is applied to the full integral image, but as this is an integral image of a subsampled copy of the main image, the processing required by the face detector may be proportionately reduced. If the subsample is ¼ of the main image, then this implies that the processing time is only 25% of that for the full image.

This approach is particularly amenable to hardware embodiments where the subsampled image memory space can be scanned by a fixed size DMA window and digital logic to implement a Haar-feature classifier chain can be applied to this DMA window. However, certain embodiment may use one or more different sizes of classifier or several sizes of classifier (e.g., in a software embodiment), or multiple fixed-size classifiers may be used (e.g., in a hardware embodiment). An advantage is that a smaller integral image is calculated.

After application of the fast face detector [280], newly detected candidate face regions [141] may be passed onto a face tracking module [111] when it is desired to use face tracking, where one or more face regions confirmed from previous analysis [145] may be merged with the new candidate face regions prior to being provided [142] to a face tracker [290].

The face tracker [290] provides a set of confirmed candidate regions [143] back to the tracking module [111]. Additional image processing filters are applied by the tracking module [111] to confirm either that these confirmed regions [143] are face regions or to

maintain regions as candidates if they have not been confirmed as such by the face tracker [290]. A final set of face regions [145] can be output by the module [111] for use elsewhere in the camera or to be stored within or in association with an acquired image for later processing either within the camera or offline; as well as to be used in the next iteration of face tracking.

After the main image acquisition chain is completed a full-size copy of the main image [130] will normally reside in the system memory [140] of the image acquisition system. This may be accessed by a candidate region extractor [125] component of the face tracker [290] which selects image patches based on candidate face region data [142] obtained from the face tracking module [111]. These image patches for each candidate region are passed to an integral image generator [115] which passes the resulting integral images to a variable-sized detector [121], as one possible example a VJ detector, which then applies a classifier chain, preferably at least a 32 classifier chain, but fewer than 32 are used in some embodiments, to the integral image for each candidate region across a range of different scales.

The range of scales [144] employed by the face detector [121] is determined and supplied by the face tracking module [111] and is based partly on statistical information relating to the history of the current candidate face regions [142] and partly on external metadata determined from other subsystems within the image acquisition system.

As an example of the former, if a candidate face region has remained consistently at a particular size for a certain number of acquired image frames then the face detector [121] may be applied at this particular scale and perhaps at one scale higher (i.e. 1.25 time larger) and one scale lower (i.e. 1.25 times lower).

As an example of the latter, if the focus of the image acquisition system has moved to infinity, then the smallest scalings would be applied in the face detector [121]. Normally these scalings would not be employed because they are applied a greater number of times to the candidate face region in order to cover it completely. The candidate face region will have a minimum size beyond which it should not decrease, and this is in order to allow for localized movement of the camera by a user between frames. In some image acquisition systems which contain motion sensors it may be possible to track such localized movements and this information may be employed to further improve the selection of scales and the size of candidate regions.

The candidate region tracker [290] provides a set of confirmed face regions [143] based on full variable size face detection of the image patches to the face tracking module [111]. Clearly, some candidate regions will have been confirmed while others will have been rejected and these can be explicitly returned by the tracker [290] or can be calculated by the tracking module [111] by analyzing the difference between the confirmed regions [143] and the candidate regions [142]. In either case, the face tracking module [111] can then apply alternative tests to candidate regions rejected by the tracker [290] (as explained below) to determine whether these should be maintained as candidate regions [142] for the next cycle of tracking or whether these should indeed be removed from tracking.

Once the set of confirmed candidate regions [145] has been determined by the face tracking module [111], the module [111] communicates with the sub-sampler [112] to determine when the next acquired image is to be sub-sampled and so provided to the detector [280] and also to provide the resolution [146] at which the next acquired image is to be sub-sampled.

It will be seen that where the detector [280] does not run when the next image is acquired, the candidate regions [142] provided to the extractor [125] for the next acquired image will be the regions [145] confirmed by the tracking module [111] from the last acquired image. On the other hand, when the face detector [280] provides a new set of candidate regions [141] to the face tracking module [111], these candidate regions are merged with the previous set of confirmed regions [145] to provide the set of candidate regions [142] to the extractor [125] for the next acquired image.

Figure 2 illustrates a exemplary workflow. The illustrated process is split into (i) a detection/initialization phase which finds new candidate face regions [141] using the fast face detector [280] which operates on a subsampled version of the full image; (ii) a secondary face detection process [290] which operates on extracted image patches for the candidate regions [142], which are determined based on the location of faces in one or more previously acquired image frames and (iii) a main tracking process which computes and stores a statistical history of confirmed face regions [143]. Although the application of the fast face detector [280] is illustrated as occurring prior to the application of the candidate region tracker [290], the order is not critical and the fast detection is not necessarily executed on every frame and in certain circumstances may be spread across multiple frames. Also, face detection may be used for various applications such as face recognition whether or not face

tracking is also used.

In step 205, the main image is acquired and in step 210 primary image processing of that main image is performed as described in relation to Figure 1. The sub-sampled image is generated by the subsampler [112] and an integral image is generated therefrom by the generator [115], step 211 as described previously. The integral image is passed to the fixed size face detector [120] and the fixed size window provides a set of candidate face regions [141] within the integral image to the face tracking module, step 220. The size of these regions is determined by the sub-sampling scale [146] specified by the face tracking module to the sub-sampler and this scale is based on the analysis of the previous sub-sampled/integral images by the detector [280] and patches from previous acquired images by the tracker [290] as well as other inputs such as camera focus and movement.

The set of candidate regions [141] is merged with the existing set of confirmed regions [145] to produce a merged set of candidate regions [142] to be provided for confirmation, step 242. For the candidate regions [142] specified by the face tracking module 111, the candidate region extractor [125] extracts the corresponding full resolution patches from an acquired image, step 225. An integral image is generated for each extracted patch, step 230 and variable-sized face detection is applied by the face detector 121 to each such integral image patch, for example, a full Viola-Jones analysis. These results [143] are in turn fed back to the face-tracking module [111], step 240.

The tracking module [111] processes these regions [143] further before a set of confirmed regions [145] is output. In this regard, additional filters can be applied by the module 111 either for regions [143] confirmed by the tracker [290] or for retaining candidate regions [142] which may not have been confirmed by the tracker 290 or picked up by the detector [280], step 245.

For example, if a face region had been tracked over a sequence of acquired images and then lost, a skin prototype could be applied to the region by the module [111] to check if a subject facing the camera had just turned away. If so, this candidate region could be maintained for checking in the next acquired image to see if the subject turns back to face the camera. Depending on the sizes of the confirmed regions being maintained at any given time and the history of their sizes, e.g. whether they are getting bigger or smaller, the module 111 determines the scale [146] for sub-sampling the next acquired image to be analyzed by the

detector [280] and provides this to the sub-sampler [112], step 250.

The fast face detector [280] need not run on every acquired image. So for example, where only a single source of sub-sampled images is available, if a camera acquires 60 frames per second, 15-25 sub-sampled frames per second (fps) may be required to be provided to the camera display for user previewing. These images are sub-sampled at the same scale and at a high enough resolution for the display. Some or all of the remaining 35-45 fps can be sampled at the scale determined by the tracking module [111] for face detection and tracking purposes.

The decision on the periodicity in which images are being selected from the stream may be based on a fixed number or alternatively be a run-time variable. In such cases, the decision on the next sampled image may be determined on the processing time it took for the previous image, in order to maintain synchronicity between the captured real-time stream and the face tracking processing. Thus in a complex image environment the sample rate may decrease.

Alternatively, the decision on the next sample may also be performed based on processing of the content of selected images. If there is no significant change in the image stream, the full face tracking process might not be performed. In such cases, although the sampling rate may be constant, the images will undergo a simple image comparison and only if it is decided that there is justifiable differences, will the face tracking algorithms be launched.

It will also be noted that the face detector [280] may run at regular or irregular intervals. So for example, if the camera focus is changed significantly, then the face detector may be run more frequently and particularly with differing scales of sub-sampled image to try to detecting faces which should be changing in size. Alternatively, where focus is changing rapidly, the detector [280] could be skipped for intervening frames, until focus has stabilised. However, it is generally only when focus goes to infinity that the highest resolution integral image is produced by the generator [115].

In this latter case, the detector in some embodiments may not be able to cover the entire area of the acquired, subsampled, image in a single frame. Accordingly the detector may be applied across only a portion of the acquired, subsampled, image on a first frame, and across the remaining portion(s) of the image on subsequent acquired image frames. In one embodiment, the detector is applied to the outer regions of the acquired image on a first

acquired image frame in order to catch small faces entering the image from its periphery, and on subsequent frames to more central regions of the image.

An alternative way of limiting the areas of an image to which the face detector 120 is to be applied comprises identifying areas of the image which include skin tones. US 6,661,907, discloses one such technique for detecting skin tones and subsequently only applying face detection in regions having a predominant skin color.

In one embodiment, skin segmentation 190 is preferably applied to the sub-sampled version of the acquired image. If the resolution of the sub-sampled version is not sufficient, then a previous image stored at image store 150 or a next sub-sampled image are preferably used when the two images are not too different in content from the current acquired image. Alternatively, skin segmentation 190 can be applied to the full size video image 130.

In any case, regions containing skin tones are identified by bounding rectangles and these bounding rectangles are provided to the integral image generator 115 which produces integral image patches corresponding to the rectangles in a manner similar to the tracker integral image generator 115.

Not alone does this approach reduce the processing overhead associated with producing the integral image and running face detection, but in certain embodiments, it also allows the face detector 120 to apply more relaxed face detection to the bounding rectangles, as there is a higher chance that these skin-tone regions do in fact contain a face. So for a VJ detector 120, a shorter classifier chain can be employed to more effectively provide similar quality results to running face detection over the whole image with longer VJ classifiers required to positively detect a face.

Further improvements to face detection are also possible. For example, it has been found that face detection is significantly dependent on illumination conditions and so small variations in illumination can cause face detection to fail, causing somewhat unstable detection behavior.

In one embodiment, confirmed face regions 145 are used to identify regions of a subsequently acquired subsampled image on which luminance correction should be performed to bring the regions of interest of the image to be analyzed to the desired parameters. One example of such correction is to improve the luminance contrast within the regions of the subsampled image defined by the confirmed face regions 145.

Contrast enhancement may be used to increase the local contrast of an image, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities for pixels of a region when represented on a histogram which would otherwise be closely distributed can be better distributed. This allows for areas of lower local contrast to gain a higher contrast without affecting the global contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values.

The method is useful in images with backgrounds and foregrounds that are both bright or both dark. In particular, the method can lead to better detail in photographs that are over or under-exposed. Alternatively, this luminance correction could be included in the computation of an "adjusted" integral image in the generators 115.

In another improvement, when face detection is being used, the camera application is set to dynamically modify the exposure from the computed default to a higher values (from frame to frame, slightly overexposing the scene) until the face detection provides a lock onto a face. In a separate embodiment, the face detector 120 will be applied to the regions that are substantively different between images. Note that prior to comparing two sampled images for change in content, a stage of registration between the images may be needed to remove the variability of changes in camera, caused by camera movement such as zoom, pan and tilt.

It is possible to obtain zoom information from camera firmware and it is also possible using software techniques which analyze images in camera memory 140 or image store 150 to determine the degree of pan or tilt of the camera from one image to another.

In one embodiment, the acquisition device is provided with a motion sensor 180, as illustrated in Figure 1, to determine the degree and direction of pan from one image to another so avoiding the processing requirement of determining camera movement in software. Motion sensors may be incorporated in digital cameras, e.g., based on accelerometers, but optionally based on gyroscopic principals, primarily for the purposes of warning or compensating for hand shake during main image capture. In this context, US patent 4,448,510, Murakoshi, discloses such a system for a conventional camera, or US patent 6,747,690, Molgaard, discloses accelerometer sensors applied within a modern digital camera.

Where a motion sensor is incorporated in a camera, it may be optimized for small movements around the optical axis. The accelerometer may incorporate a sensing module which generates a signal based on the acceleration experienced and an amplifier module which determines the range of accelerations which can effectively be measured. The accelerometers may allow software control of the amplifier stage which allows the sensitivity to be adjusted.

The motion sensor 180 could equally be implemented with MEMS sensors of the sort which will be incorporated in next generation consumer cameras and camera-phones. In any case, when the camera is operable in face tracking mode, i.e. constant video acquisition as distinct from acquiring a main image, shake compensation might not be used because image quality is lower. This provides the opportunity to configure the motion sensor 180, to sense large movements, by setting the motion sensor amplifier module to low gain. The size and direction of movement detected by the sensor 180 is provided to the face tracker 111. The approximate size of faces being tracked is already known and this enables an estimate of the distance of each face from the camera. Accordingly, knowing the approximate size of the large movement from the sensor 180 allows the approximate displacement of each candidate face region to be determined, even if they are at differing distances from the camera.

Thus, when a large movement is detected, the face tracker 111 shifts the location of candidate regions as a function of the direction and size of the movement. Alternatively, the size of the region over which the tracking algorithms are applied may also be enlarged (and, if necessary, the sophistication of the tracker may be decreased to compensate for scanning a larger image area) as a function of the direction and size of the movement.

When the camera is actuated to capture a main image, or when it exits face tracking mode for any other reason, the amplifier gain of the motion sensor 180 is returned to normal, allowing the main image acquisition chain 105,110 for full-sized images to employ normal shake compensation algorithms based on information from the motion sensor 180.In alternative embodiments, sub-sampled preview images for the camera display can be fed through a separate pipe than the images being fed to and supplied from the image sub-sampler [112] and so every acquired image and its sub-sampled copies can be available both to the detector [280] as well as for camera display.

In addition to periodically acquiring samples from a video stream, the process may also be applied to a single still image acquired by a digital camera. In this case, the stream for the

face tracking comprises a stream of preview images and the final image in the series is the full resolution acquired image. In such a case, the face tracking information can be verified for the final image in a similar fashion to that illustrated in Figure 2. In addition, the information such as coordinates or mask of the face may be stored with the final image. Such data for example may fit as an entry in the saved image header, for future post processing, whether in the acquisition device or at a later stage by an external device.

Figures 3A-3D illustrate operations of certain embodiments through worked examples. Figure 3A illustrates the result at the end of a detection & tracking cycle on a frame of video or a still within a series of stills, and two confirmed face regions [301, 302] of different scales are shown. In this embodiment, for pragmatic reasons, each face region has a rectangular bounding box, as it is easier to make computations on rectangular regions. This information is recorded and output as [145] by the tracking module [111] of Figure 1. Based on the history of the face regions [301,302], the tracking module [111] may decide to run fast face tracking with a classifier window of the size of face region [301] with an integral image being provided and analyzed accordingly.

Figure 3B illustrates the situation after the next frame in a video sequence is captured and the fast face detector has been applied to the new image. Both faces have moved [311, 312] and are shown relative to previous face regions [301, 302]. A third face region [303] has appeared and has been detected by the fast face detector [303]. In addition the fast face detector has found the smaller of the two previously confirmed faces [304] because it is at the correct scale for the fast face detector. Regions [303] and [304] are supplied as candidate regions [141] to the tracking module [111]. The tracking module merges this new candidate region information [141], with the previous confirmed region information [145] comprising regions [301] [302] to provide a set of candidate regions comprising regions [303], [304] and [302] to the candidate region extractor [290]. The tracking module [111] knows that the region [302] has not been picked up by the detector [280]. This may be because the face has disappeared, remains at a size that could not have been detected by the detector [280] or has changed size to a size that could not have been detected by the detector [280]. Thus, for this region, the module [111] will specify a large patch [305].

The large patch 305 may be as illustrated at Figure 3C around the region [302] to be checked by the tracker [290]. Only the region [303] bounding the newly detected face candidate needs

to be checked by the tracker [290], whereas because the face [301] is moving a relatively large patch [306] surrounding this region is specified to the tracker [290].

Figure 3C illustrates the situation after the candidate region extractor operates upon the image, candidate regions [306, 305] around both of the confirmed face regions [301, 302] from the previous video frame as well as new region [303] are extracted from the full resolution image [130]. The size of these candidate regions has been calculated by the face tracking module [111] based partly on partly on statistical information relating to the history of the current face candidate and partly on external metadata determined from other subsystems within the image acquisition system. These extracted candidate regions are now passed on to the variable sized face detector [121] which applies a VJ face detector to the candidate region over a range of scales. The locations of one or more confirmed face regions, if any, are then passed back to the face tracking module [111].

Figure 3D illustrates the situation after the face tracking module [111] has merged the results from both the fast face detector [280] and the face tracker [290] and applied various confirmation filters to the confirmed face regions. Three confirmed face regions have been detected [307, 308, 309] within the patches [305, 306, 303]. The largest region [307] was known but had moved from the previous video frame and relevant data is added to the history of that face region. The other previously known region [308] which had moved was also detected by the fast face detector which serves as a double-confirmation and these data are added to its history. Finally, a new face region [303] was detected and confirmed and a new face region history must be initiated for this newly detected face. These three face regions are used to provide a set of confirmed face regions [145] for the next cycle.

There are many possible applications for the regions 145 supplied by the face tracking module. For example, the bounding boxes for each of the regions [145] can be superimposed on the camera display to indicate that the camera is automatically tracking detected face(s) in a scene. This can be used for improving various pre-capture parameters. One example is exposure, ensuring that the faces are well exposed. Another example is auto-focusing, by ensuring that focus is set on a detected face or indeed to adjust other capture settings for the optimal representation of the face in an image.

The corrections may be done as part of the pre-processing adjustments. The location of the face tracking may also be used for post processing and in particular selective post processing

where the regions with the faces may be enhanced. Such examples include sharpening, enhancing saturation, brightening or increasing local contrast. The preprocessing using the location of faces may also be used on the regions without the face to reduce their visual importance, for example through selective blurring, de-saturation, or darkening.

Where several face regions are being tracked, then the longest lived or largest face can be used for focusing and can be highlighted as such. Also, the regions [145] can be used to limit the areas on which for example red-eye processing is performed when required. Other post-processing which can be used in conjunction with the light-weight face detection described above is face recognition. In particular, such an approach can be useful when combined with more robust face detection and recognition either running on the same or an off-line device that has sufficient resources to run more resource consuming algorithms.

In this case, the face tracking module [111] reports the location of any confirmed face regions [145] to the in-camera firmware, preferably together with a confidence factor. When the confidence factor is sufficiently high for a region, indicating that at least one face is in fact present in an image frame, the camera firmware runs a light-weight face recognition algorithm [160] at the location of the face, for example a DCT-based algorithm. The face recognition algorithm [160] uses a database [161] preferably stored on the camera comprising personal identifiers and their associated face parameters.

In operation, the module [160] collects identifiers over a series of frames. When the identifiers of a detected face tracked over a number of preview frames are predominantly of one particular person, that person is deemed by the recognition module to be present in the image. One or both of the identifier of the person and the last known location of the face are stored either in the image (in a header) or in a separate file stored on the camera storage [150]. This storing of the person's ID can occur even when the recognition module [160] has failed for the immediately previous number of frames but for which a face region was still detected and tracked by the module [111].

When an image is copied from camera storage to a display or permanent storage device such as a PC (not shown), the person ID's are copied along with the images. Such devices are generally more capable of running a more robust face detection and recognition algorithm and then combining the results with the recognition results from the camera, giving more weight to recognition results from the robust face recognition (if any). The combined

identification results are presented to the user, or if identification was not possible, the user is asked to enter the name of the person that was found. When the user rejects an identification or a new name is entered, the PC retrains its face print database and downloads the appropriate changes to the capture device for storage in the light-weight database [161]. When multiple confirmed face regions [145] are detected, the recognition module [160] can detect and recognize multiple persons in the image.

It is possible to introduce a mode in the camera that does not take a shot until persons are recognized or until it is clear that persons are not present in the face print database, or alternatively displays an appropriate indicator when the persons have been recognized. This allows reliable identification of persons in the image.

This feature solves the problem where algorithms using a single image for face detection and recognition may have lower probability of performing correctly. In one example, for recognition, if the face is not aligned within certain strict limits it is not possible to accurately recognize a person. This method uses a series of preview frames for this purpose as it can be expected that a reliable face recognition can be done when many more variations of slightly different samples are available.

Further improvements to the efficiency of systems described herein are possible. For example, a face detection algorithm may employ methods or use classifiers to detect faces in a picture at different orientations: 0, 90, 180 and 270 degrees. According to a further embodiment, the camera is equipped with an orientation sensor. This can comprise a hardware sensor for determining whether the camera is being held upright, inverted or tilted clockwise or anti-clockwise. Alternatively, the orientation sensor can comprise an image analysis module connected either to the image acquisition hardware 105, 110 or camera memory 140 or image store 150, each as illustrated in Figure 1, for quickly determining whether images are being acquired in portrait or landscape mode and whether the camera is tilted clockwise or anti-clockwise.

Once this determination is made, the camera orientation can be fed to one or both of the face detectors 120, 121. The detectors need then only apply face detection according to the likely orientation of faces in an image acquired with the determined camera orientation. This feature significantly reduces face detection processing overhead, for example, by avoiding the employing of classifiers which are unlikely to detect faces or increase its accuracy by running

classifiers more likely to detects faces in a given orientation more often.

According to another embodiment, there is provided a method for image recognition in a collection of digital images that includes training image classifiers and retrieving a sub-set of images from the collection. The training of the image classifiers preferably includes one, more than one or all of the following: For each image in the collection, any regions within the image that correspond to a face are identified. For each face region and any associated peripheral region, feature vectors are determined for each of the image classifiers. The feature vectors are stored in association with data relating to the associated face region.

The retrieval of the sub-set of images from the collection preferably includes one, more than one or all of the following: At least one reference region including a face to be recognized is/are selected from an image. At least one classifier on which said retrieval is to be based is/are selected from the image classifiers. A respective feature vector for each selected classifier is determined for the reference region. The sub-set of images is retrieved from within the image collection in accordance with the distance between the feature vectors determined for the reference region and the feature vectors for face regions of the image collection.

A component for image recognition in a collection of digital images is further provided including a training module for training image classifiers and a retrieval module for retrieving a sub-set of images from the collection.

The training module is preferably configured according to one, more than one or all of the following: For each image in the collection, any regions are identified that correspond to a face in the image. For each face region and any associated peripheral region, feature vectors are determined for each of the image classifiers. The feature vectors are stored in association with data relating to the associated face region.

The retrieval module is preferably configured according to one, more than one or all of the following: At least one reference region including a face to be recognized is/are selected from an image. At least one image classifier is/are selected on which the retrieval is to be based. A respective feature vector is determined for each selected classifier of the reference region. A sub-set of images is selected from within the image collection in accordance with

the distance between the feature vectors determined for the reference region and the feature vectors for face regions of the image collection.

In a further aspect there is provided a corresponding component for image recognition. In this embodiment, the training process cycles automatically through each image in an image collection, employing a *face detector* to determine the location of face regions within an image. It then extracts and normalizes these regions and associated non-face peripheral regions which are indicative of, for example, the hair, clothing and/or pose of the person associated with the determined face region(s). Initial training data is used to determine a basis vector set for each face classifier.

A basis vector set comprises a selected set of attributes and reference values for these attributes for a particular classifier. For example, for a DCT classifier, a basis vector could comprise a selected set of frequencies by which selected image regions are best characterized for future matching and/or discrimination and a reference value for each frequency. For other classifiers, the reference value can simply be the origin (zero value) within a vector space.

Next, for each determined, extracted and normalized face region, at least one feature vector is generated for at least one face-region based classifier and where an associated non-face region is available, at least one further feature vector is generated for a respective non-face region based classifier. A feature vector can be thought of as an identified region's coordinates within the basis vector space relative to the reference value.

These data are then associated with the relevant image and face/peripheral region and are stored for future reference. In this embodiment, image retrieval may either employ a user selected face region or may automatically determine and select face regions in a newly acquired image for comparing with other face regions within the selected image collection. Once at least one face region has been selected, the retrieval process determines (or if the image was previously "trained", loads) feature vectors associated with at least one face-based classifier and at least one non-face based classifier. A comparison between the selected face region and all other face regions in the current image collection will next yield a set of distance measures for each classifier. Further, while calculating this set of distance measures, mean and variance values associated with the statistical distribution of the distance measures for each classifier are calculated. Finally these distance measures are preferably normalized

using the mean and variance data for each classifier and are summed to provide a combined distance measure which is used to generate a final ranked similarity list.

In another embodiment, the classifiers include a combination of wavelet domain PCA (principle component analysis) classifier and 2D-DCT (discrete cosine transform) classifier for recognizing face regions. These classifiers do not require a training stage for each new image that is added to an image collection. For example, techniques such as ICA (independent component analysis) or the Fisher Face technique which employs LDA (linear discriminant analysis) are well known face recognition techniques which adjust the basis vectors during a training stage to cluster similar images and optimize the separation of these clusters.

The combination of these classifiers is robust to different changes in face poses, illumination, face expression and image quality and focus (sharpness). PCA (principle component analysis) is also known as the eigenface method. A summary of conventional techniques that utilize this method is found in *Eigenfaces for Recognition,* Journal of Cognitive Neuroscience, 3(1), 1991 to Turk et al. This method is sensitive to facial expression, small degrees of rotation and different illuminations. In the preferred embodiment, high frequency components from the image that are responsible for slight changes in face appearance are filtered. Features obtained from low pass filtered sub-bands from the wavelet decomposition are significantly more robust to facial expression, small degrees of rotation and different illuminations than conventional PCA.

In general, the steps involved in implementing the PCA/Wavelet technique include: (i) the extracted, normalized face region is transformed into gray scale; (ii) wavelet decomposition in applied using Daubechie wavelets; (iii) histogram equalization is performed on the grayscale LL sub-band representation; next, (iv) the mean LL sub-band is calculated and subtracted from all faces and (v) the 1st level LL sub-band is used for calculating the covariance matrix and the principal components (eigenvectors). The resulting eigenvectors (basis vector set) and the mean face are stored in a file after training so they can be used in determining the principal components for the feature vectors for detected face regions. Alternative embodiments may be discerned from the discussion in H. Lai, P. C. Yuen, and G. C. Feng, "Face recognition using holistic Fourier invariant features" Pattern Recognition, vol. 34, pp. 95–109, 2001.

In the 2D Discrete Cosine Transform classifier, the spectrum for the DCT transform of the face region can be further processed to obtain more robustness (see also, *Application of the DCT Energy Histogram for Face Recognition*, in Proceedings of the 2nd International Conference on Information Technology for Application (ICITA 2004) to Tjahyadi et al.

The steps involved in this technique are generally as follows: (i) the resized face is transformed to an indexed image using a 256 color gif colormap; (ii) the 2D DCT transform is applied; (iii) the resulting spectrum is used for classification; (iv) for comparing similarity between DCT spectra the Euclidian distance was used. Examples of non-face based classifiers are based on color histogram, color moment, colour correlogram, banded colour correlogram, and wavelet texture analysis techniques. An implementaton of color histogram is described in "CBIR method based on color-spatial feature," *IEEE Region 10th Ann. Int. Conf. 1999* (*TENCON'99*, Cheju, Korea, 1999). Use of the colour histogram is, however, typically restricted to classification based on the color information contained within one or more sub-regions of the image.

Color moment may be used to avoid the quantization effects which are found when using the color histogram as a classifier (see also "Similarity of color images," *SPIE Proc. pp. 2420* (1995) to Stricker *et al.*). The first three moments (mean, standard deviation and skews) are extracted from the three color channels and therefore form a 9-dimensional feature vector.

The color auto-correlogram (see, US 6,246,790 to Huang et al.) provides an image analysis technique that is based on a three-dimensional table indexed by color and distance between pixels which expresses how the spatial correlation of color changes with distance in a stored image. The color correlogram may be used to distinguish an image from other images in a database. It is effective in combining the color and texture features together in a single classifier (see also, "Image indexing using color correlograms," In *IEEE Conf. Computer Vision and Pattern Recognition, PP. 762 et seq* (1997) to Huang et al.).

In certain embodiments, the color correlogram is implemented by transforming the image from RGB color space, and reducing the image colour map using dithering techniques based on minimum variance quantization. Variations and alternative embodiments may be discerned from *Variance based color image quantization for frame buffer display," Color Res. Applicat., vol.15, no. 1, pp. 52–58, 1990 to by Wan et al.* Reduced colour maps of 16, 64, 256 colors are achievable. For 16 colors the VGA colormap may be used and for 64 and

256 colors, a gif colormap may be used. A maximum distance set $D = 1; 3; 5; 7$ may be used for computing auto-correlogram to build a $N \times D$ dimension feature vector where N is the number of colors and D is the maximum distance.

The color autocorrelogram and banded correlogram may be calculated using a fast algorithm (see, e.g., "Image Indexing Using Color Correlograms" from the Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97) to Huang et al.). Wavelet texture analysis techniques (see, e.g., "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Processing* **2**(4), 429 (1993) to Chang et al.) may also be advantageously used. In order to extract the wavelet based texture, the original image is decomposed into 10 de-correlated sub-bands through 3-level wavelet transform. In each sub-band, the standard deviation of the wavelet coefficients is extracted, resulting in a 10-dimensional feature vector.

Another embodiment is described in relation to Figure 4. This takes the form of a set of software modules 1162 implemented on a desktop computer 1150. A second preferred embodiment provides an implementation within an embedded imaging appliance such as a digital camera.

In this embodiment, a program may be employed in a desktop computer environment and may either be run as a stand-alone program, or alternatively, may be integrated in existing applications or operating system (OS) system components to improve their functionality.

IMAGE ANALYSIS MODULE

An image analysis module 1156, such as that illustrated at Figure 4, cycles through a set of images 1170-1...1180-2 and determines, extracts, normalizes and analyzes face regions and associated peripheral regions to determine feature vectors for a plurality of face and non-face classifiers. The module then records this extracted information in an image data set record. Components of this module are also used in both training and sorting/retrieval modes of the embodiment. The module is called from a higher level workflow and in its normal mode of usage is passed a set of images which, as illustrated at Figure 7, are analyzed [2202]. The module loads/acquires the next image [2202] and detects any face regions in said image [2204]. If no face regions were found, then flags in the image data record for that image are updated to indicate that no face regions were found. If the current image is not the last image

in the image set being analyzed [2208], upon image subsampling [2232], face and peripheral region extraction [2206] and region normalization [2207], the next image is loaded/acquired [2204]. If this was the last image, then the module will exit to a calling module. Where at least one face region is detected the module next extracts and normalizes each detected face region and, where possible, any associated peripheral regions.

Face region normalization techniques can range from a simple re-sizing of a face region to more sophisticated 2D rotational and affine transformation techniques and to highly sophisticated 3D face modeling methods.

IMAGE SORTING/RETRIEVAL PROCESS

The workflow for an image sorting/retrieval process or module is illustrated at Figures 5 and 6A-6F and is initiated from an image selection or acquisition process (see US 2006/0140455) as the final process step [1140]. It is assumed that when the image sorting/retrieval module is activated [1140] it will also be provided with at least two input parameters providing access to (i) the image to be used for determining the search/sort/classification criteria, and (ii) the image collection data set against which the search is to be performed. If a data record is determined to not be available [1306] and has not already been determined for the search image which proceeds to select persons and search criteria in the image [1308], then main image analysis module is next applied to it to generate this data record [1200]. The image is next displayed to a user who may be provided options to make certain selections of face regions to be used for searching and/or also of the classifiers to be used in the search [1308]. Alternatively, the search criteria may be predetermined or otherwise automated through a configuration file and step [1308] may thus be automatic. User interface aspects are described in detail at US 2006/0140455.

After a reference region comprising the face and/or peripheral regions to be used in the retrieval process is selected (or determined automatically) the main retrieval process is initiated [1310] either by user interaction or automatically in the case where search criteria are determined automatically from a configuration file. The main retrieval process is described in step [1312] and comprises three main sub-processes which are iteratively performed for each classifier to be used in the sorting/retrieval process:

(i) Distances are calculated in the current classifier space between the feature vector

for the reference region and corresponding feature vector(s) for the face/peripheral regions for all images in the image collection to be searched [1312-1]. In the preferred embodiment, the Euclidean distance is used to calculate these distances which serve as a measure of similarity between the reference region and face/peripheral regions in the image collection.

(ii) The statistical mean and standard deviation of the distribution of these calculated distances is determined and stored temporarily [1312-2].

(iii) The determined distances between the reference region and the face/peripheral regions in the image collection are next normalized [1312-3] using the mean and standard deviation determined in step [1312-2].

These normalized data sets may now be combined in a decision fusion process [1314] which generates a ranked output list of images. These may then be displayed by a UI module [1316].

An additional perspective on the process steps [1312-1, 1312-2 and 1312-3] is given in US 2006/0140455. The classifier space [1500] for a classifier may be such as the Wavelet/PCA face recognition described at US 2006/0140455. The basis vector set, $[\lambda_1, \lambda_2, \ldots, \lambda_n]$ may be used to determine feature vectors for this classifier. The average or mean face is calculated [1501] during the training phase and its vector position [1507] in classifier space [1500] is subtracted from the absolute position of all face regions. Thus, exemplary face regions [1504-1a, 1504-2a and 1504-3a] have their positions [1504-1b, 1504-2b and 1504-3b] in classifier space defined in vector terms relative to the mean face [1501].

After a particular face region [1504-2a] is selected by the user [1308] the distances to all other face regions within a particular image collection are calculated. The face regions [1504-1a] and [1504-3a] are shown as illustrative examples. The associated distances (or non-normalized rankings) are given as [1504-1c] and [1504-3c].

An analogous case arises when the distances in classifier space are measured in absolute terms from the origin, rather than being measured relative to the position of an averaged, or mean face. For example, the color correlogram technique as used in certain embodiments is a classifier of this type which does not have the equivalent of a mean face.

The distances from the feature vector for the reference region [1504-2a] and [1509-2a] to the feature vectors for all other face regions may be calculated in a number of ways. In one embodiment, Euclidean distance is used, but other distance metrics may be advantageously employed for certain classifiers other than those described here.

METHODS FOR COMBINING CLASSIFIER SIMILARITY MEASURES

STATISTICAL NORMALIZATION METHOD

A technique is preferably used for normalizing and combining the multiple classifiers to reach a final similarity ranking. The process may involve a set of multiple classifiers, C_1, C_2 ...C_N and may be based on a statistical determination of the distribution of the distances of all patterns relevant to the current classifier (face or peripheral regions in our embodiment) from the selected reference region. For most classifiers, this statistical analysis typically yields a normal distribution with a mean value M_{Cn} and a variance V_{Cn}.

IN-CAMERA IMPLEMENTATION

As imaging appliances continue to increase in computing power, memory and non-volatile storage, it will be evident to those skilled in the art of digital camera design that many advantages can be provided as an in-camera image sorting sub-system. An exemplary embodiment is illustrated in Figure 7.

Following the main image acquisition process [2202] a copy of the acquired image is saved to the main image collection [2212] which will typically be stored on a removable compact-flash or multimedia data card [2214]. The acquired image may also be passed to an image subsampler [2232] which generates an optimized subsampled copy of the main image and stores it in a subsampled image collection [2216]. These subsampled images may advantageously be employed in the analysis of the acquired image.

The acquired image (or a subsampled copy thereof) is also passed to a face detector module [2204] followed by a face and peripheral region extraction module [2206] and a region normalization module [2207]. The extracted, normalized regions are next passed to the main image analysis module [2208] which generates an image data record [1409] for the current image. The main image analysis module may also be called from the training module [2230] and the image sorting/retrieval module [2218].

A UI module [2220] facilitates the browsing & selection of images [2222], the selection of one or more face regions [2224] to use in the sorting/retrieval process [2218]. In addition classifiers may be selected and combined [2226] from the UI Module [2220].

Various combinations are possible where certain modules are implemented in a digital camera and others are implemented on a desktop computer.

ILLUMINATION CLASSIFIERS

A branched classifier chain may be used for simultaneous classification of faces and classification of uneven (or even) illumination. In certain embodiments, a classifier chain is constructed that, after an initial set of feature detectors that reject the large majority of objects within an image as non-faces, applies a set of, for example 3, 4, 5, 6, 7, 8 or 9, feature detectors. The feature detectors may tuned so that they accept faces that are illuminated from the top, bottom and left or right illumination (due to faces being left-right symmetrical), OR top, bottom, left or right, and even illumination, OR top, bottom, left, right and even illumination, OR top, left, right, bottom, bottom-right, bottom-left, top-right, and top-left illumination, OR top, left, right, bottom, top right, top left, bottom right, bottom left and even illumination, OR top, bottom, right or left or both, top-right or top-left or both, bottom-right or bottom-left or both, and even. Other combinations are possible, and some may be excluded, e.g., after application of one classifier provides a determination that a face exists within the image or a sub-window of the image of a certain illumination. When one of the classifier branches accepts the face, it can be said that the face and the illumination of the face are detected. This detection can be used to process the image with greater attention to faces than non-faces, and/or to correct the uneven illumination condition, improving face recognition results.

Alternatively, the detected illumination problems in one detection frame may be corrected in the next frame so the face detection algorithm has a better chance of finding the face. The illumination detection comes essentially for free as the length of the classifier chain is not longer than in the previous design.

Figure 8 illustrates a face illumination normalization method in accordance with certain embodiments. A digital image is acquired at 602. One or more uneven illumination classifier sets are applied to the data at 604, beginning with one cascade at a time. The sets

may be used to find faces and/or to determine an uneven (or even) illumination condition within already detected face image. Depending on the data retrieved in 604, method according to different embodiments would next identify a face within the image at 606, or determine an uneven (or even) illumination condition for a face at 608, or both 606 and 608 contemporaneously or one after the other in either order. For example, a face may be found and then an illumination condition found for the face, or an illumination condition for an object may be found followed by a determination whether the object is a face.

It may also be determined that no single illumination condition exists at 618. If a face is determined to exist at 606, then at 616, a set of feature detector programs may be applied to reject non-face data from being identified as a face (or accept face data as being identified as a face).

If an uneven illumination condition is determined at 608, then at 610 the uneven illumination condition may be corrected for the image and/or for another image in a series of images. For example, the original image may be a preview image, and a full resolution image may be corrected either during acquisition (e.g., by adjusting a flash condition or by providing suggestions to the camera-user to move before taking the picture, etc.) or after acquisition either in-camera before or after storing a permanent image, or on an external device later-on. Corrected face image data may be generated at 612 appearing to have more uniform illumination, and the corrected face image may be stored, transmitted, applied to a face recognition program, edited and/or displayed at 614.

If it is determined at 618 that no single illumination condition applies, then the face data may be rejected or not rejected as a face at 620. If the face data is not rejected as a face at 620, then at 622, combinations of two or more classifier sets may be applied to the data.

Figures 9A-9B illustrate face detection methods in accordance with certain further embodiments. A digital image is acquired at 702. A sub-window is extracted from the image at 704. Two or more shortened face detected classifier cascades are applied to the sub-window at 706. These cascades are trained to be selectively sensitive to a characteristic of a face region.

At 708, a probability is determined that a face with a certain form of the characteristic is present within the sub-window. The characteristic may include an illumination condition, or

a pose or direction of the face relative to the camera, or another characteristic such as resolution, size, location, motion, blurriness, facial expression, blink condition, red, gold or white eye condition, occlusion condition or an appearance, e.g., of a face within a collection having multiple appearances such as shaven or unshaven, a hair style, or wearing certain jewelry, among other features. An extended face detection classifier cascade is applied at 710 for sensitivity to the form of the characteristic. A final determination is provided at 712 whether a face exists within the sub-window. If so, then optionally at 714, an uneven illumination condition for the face image may be corrected within the image and/or within a different image in a series of images. In addition, the process may return to 704 to extract a further sub-window, if any, from the image.

At 742, a digital image may be acquired, and a sub-window extracted therefrom at 744. Tow or more shortened face detection classifier cascades may be applied at 746 that are trained to be selectively sensitive to directional face illumination. A probability is determined that a face having a certain directional facial illumination condition is present within the sub-window at 748. An extended face detection classifier cascade is applied at 750 that is trained for sensitivity to the certain form of directional face illumination, e.g., top, bottom, right, left, top-right or top-left, bottom-right or bottom-left, and/or even. A final determination is provided at 752 whether a face exists within the image sub-window. A further sub-window, if any, may then be extracted by returning the process to 744 and/or an uneven illumination condition of the face may be corrected within the image and/or a different image in a series of images at 754.

The "Chain Branching" idea for Luminance is fairly straight-forward to implement and to test since it requires no alterations to the training algorithm. The variations/"mutations" of a face are considered as distinct objects and each one receives a distinct detector/cascade of classifiers. The detectors are all the same, linear chains of full extent.

In detection the straightforward approach would be to exhaustively run all the detectors and see which ones accept the window and then choose the best score. This means that the correct detector is selected at the end. However, this is not what we tested, being very time-consuming.

Chain1 = cls11 + cls12 + ... + cls1M

...

ChainN = clsN1 + clsN2 + ... + clsNM

The detectors may be run in series or in parallel or some combination thereof, and an at least partial confidence may be accumulated, viz:

Partial1 = cls11 + cls12 + ... + cls1P

...

PartialN = clsN1 + clsN2 + ... + clsNP, with P < M

The best detector is chosen at this point with maximum Partial confidence value. Only that detector continues execution with:

ChainMax = PartialMax + clsMax(P+1) + clsMax(P+2) + ... + clsMaxM

So an exemplary workflow is:

Partial1 ---

... \

PartialMax --- (choose Max) -> continue with the rest of Max

... /

PartialN ---

This approach may be applied for face pose variation and/or an illumination condition or other characteristic. In the illumination case, one may use any combination of (i) frontally illuminated faces; (ii) faces illuminated from the top; (iii) faces illuminated from bottom; (iv) faces illuminated form the left and (v) faces illuminated from right. Because of the symmetric nature of faces, one could use just one of (iv) and (v) as there is symmetry between the classifiers obtained. The training images used for determining these classifier sets may be generated using an AAM model with one parameter trained to correspond to the level of top/bottom illumination and a second parameter trained to correspond to left/right illumination.

Figures 10A-10B illustrate an exemplary detailed workflow. At 802, a sub-window is tested with a frontally illuminated partial classifier set (e.g., using 3-5 classifiers). If a cumulative probability is determined at 804 to be above a first threshold, then the face is determined to be frontally illuminated at 806, and the process is continued with this full classifier chain. If the cumulative probability is determined to be below a second threshold (which is even lower than the first threshold), then at 812 the sub-window is determined to not contain a face, and the process is returned via 864 to 802. If the cumulative probability is determined at 808 to be above a second threshold, yet below the first threshold of 804, then the sub-window is deemed to still likely be a face at 810, but not a frontally illuminated one. Thus, a next illumination specific partial classifier set is applied at 814.

The classifier can be applied in any order, although at step 814, the sub-window is tested with a top illuminated partial classifier set (e.g., using 3-5 classifiers). If the cumulative probability is determined to be above a first threshold at 816, then face is determined to be top illuminated at 818, and the process is continued with this full classifier chain. If the cumulative probability is deemed to be between the first threshold and a lower second threshold at 820, then at 822 the sub-window is determined to still likely contain a face, but not a top illuminated one, and so the process moves to 826 for applying a next illumination specific partial classifier set. If the cumulative probability is deemed to be less than the second threshold, then at 824 the sub-window is determined to not contain a face, and the process moves back through 864 to the next sub-window and 802.

At 826, a test of the sub-window is performed with a bottom illuminated partial classifier set (e.g., using 3-5 classifiers). If the cumulative probability is determined at 828 to be above a first threshold, then the face is determined to be top illuminated and at 830 the process is continued with this full classifier chain. If cumulative probability is below the first threshold, but above a lower second threshold at 832, then the sub-window is determined to still likely contain a face at 834, although not a bottom illuminated one, and so the process moves to 838 and Figure 10B to apply a next illumination specific partial classifier set. If the cumulative probability is below this second threshold though, then it is determined at 836 than the sub-window does not contain a face, and the process moves through 864 back to 802 and an next sub-window. As the sub-window had not been rejected at 810 nor 822, a further check may be performed prior to rejecting the sub-window at 836, and the same would apply at 824, as well as 846 and 858 of Figure 10B.

At 838, a test of the sub-window is performed with a left-illuminated partial classifier set (e.g., using 3-5 classifiers). If cumulative probability is deemed to be above a first threshold at 840, then the face is determined to be top illuminated, and at 842, the process is continued with this full classifier chain. Otherwise, if the cumulative probability is still deemed to be above a second threshold below the first at 844, then it is determined at 846 that the sub-window of image data is still likely to contain a face, although not a left illuminated one, and so the next illumination specific partial classifier set is applied at 850. If the cumulative probability is below the second threshold, then at 848, the sub-window is deemed to not contain a face, and so the process is moved to the next image window through 864 back to 802 at Figure 10A.

At 850, a test of the sub-window is performed with a right-illuminated partial classifier set (e.g., using 3-5 classifiers). If the cumulative probability is deemed to be above a first threshold at 852, then at 854, the sub-window is determined to contain a face that is top illuminated, and the process is continued with this full classifier chain. If at 852, however, the cumulative probability deemed to be below the first threshold, but at 856 it is deemed to be above a second threshold lower than the first, then the sub-window is still deemed to be likely to contain a face at 858, although not a right illuminated one, and so now pairs of specific partial classifier sets are applied at 862. This is because at this point, the window has not passed any of the illumination specific classifiers at their first threshold but neither has it been rejected as a face. Thus, a likely scenario is that the sub-window contains a face that is represented by a combination of illumination types. So, the two highest probability thresholds may be first applied to determine whether is it is top/bottom and/or right/left illuminated, then both full classifier sets are applied to determined if it survives as a face region. If at 856 the cumulative probability is deemed to be below the second threshold, then at 860, the sub-window is deemed not to contain a face and the processes moves through 864 to 802 to the next image sub-window.

In addition, in methods that may be performed according to preferred embodiments herein and that may have been described above, the operations have been described in selected typographical sequences. However, the sequences have been selected and so ordered for typographical convenience and are not intended to imply any particular order for performing the operations, except for those where a particular order may be expressly set forth or where those of ordinary skill in the art may deem a particular order to be necessary.

Claims:

1. A face illumination normalization method, comprising:

(a) acquiring a digital image including data corresponding to a face that appears to be illuminated unevenly;

(b) applying one or more uneven illumination classifier programs to the face data;

(c) identifying the face data as corresponding to said face within the digital image;

(d) determining an uneven illumination condition for the face also as a result of the applying of the one or more uneven illumination classifier programs;

(e) correcting the uneven illumination condition of the face based on the determining to thereby generate a corrected face image appearing to have more uniform illumination; and

(f) electronically storing, transmitting, applying a face recognition program to, editing, or displaying the corrected face image, or combinations thereof.

2. The method of claim 1, further comprising applying a face recognition program to the corrected face image.

3. The method of claim 1, wherein the detecting of the face and the determining of the uneven illumination condition of the face are performed simultaneously.

4. The method of claim 1, further comprising applying a set of feature detector programs to reject non-face data from being identified as face data.

5. The method of claim 1, further comprising applying a front illumination classifier program to the face data.

6. The method of claim 5, further comprising determining an illumination condition based on acceptance of the face data by one of the classifier programs.

7. The method of claim 6, wherein the digital image is one of multiple images in a series that include said face, and wherein said correcting is applied to a different image in the series than said digital image within which the illuminating condition is determined.

8. The method of claim 1, wherein said uneven illumination classifier programs comprise a top illumination classifier, a bottom illumination classifier, and one or both of right and left illumination classifiers.

9. The method of claim 8, further comprising applying a front illumination classifier program to the face data.

10. The method of claim 1, wherein the applying comprises applying at least two full classifier sets after determining that no single illumination condition applies and that the face data is not rejected as a face.

11. A face detection method, comprising:

(a) acquiring a digital image

(b) extracting a sub-window from said image

(c) applying two or more shortened face detection classifier cascades, trained to be selectively sensitive to a characteristic of a face region,

(d) based on the applying, determining a probability that a face with a certain form of the characteristic is present within the sub-window;

(e) based on the determining, applying an extended face detection classifier cascade trained for sensitivity to said form of said characteristic;

(f) providing a final determination that a face exists within the image sub-window; and

(g) repeating steps (b)-(e) one or more times for one or more further sub-windows from the image or one or more further characteristics, or both.

12. The method of claim 11, wherein the characteristic or characteristics comprise a directional illumination of the face region, an in-plane rotation of the face region, a 3D pose variation of the face region. a degree of smile, a degree of eye-blinking, a degree of eye-winking, a degree of mouth opening, facial blurring, eye-defect, facial shadowing, facial occlusion, facial color, or facial shape, or combinations thereof.

13. The method of claim 11, wherein the characteristic comprises a directional illumination, and the method further comprises determining an uneven illumination condition by applying one or more uneven illumination classifier cascades.

14. The method of claim 13, further comprising applying a front illumination classifier cascade.

15. The method of claim 14, further comprising determining an illumination condition of a face within a sub-window based on acceptance by one of the classifier cascades.

16. The method of claim 15, wherein the digital image is one of multiple images in a series that include the face, and the method further comprises correcting an uneven illumination condition of the face within a different image in the series than said digital image within which the illuminating condition is determined.

17. The method of claim 13, wherein said uneven illumination classifier cascades comprise a top illumination classifier, a bottom illumination classifier, and one or both of right and left illumination classifiers.

18. A digital image acquisition device including an optoelectonic system for acquiring a digital image, and a digital memory having stored therein processor-readable code for programming the processor to perform a face illumination normalization method as in any of the previous claims.

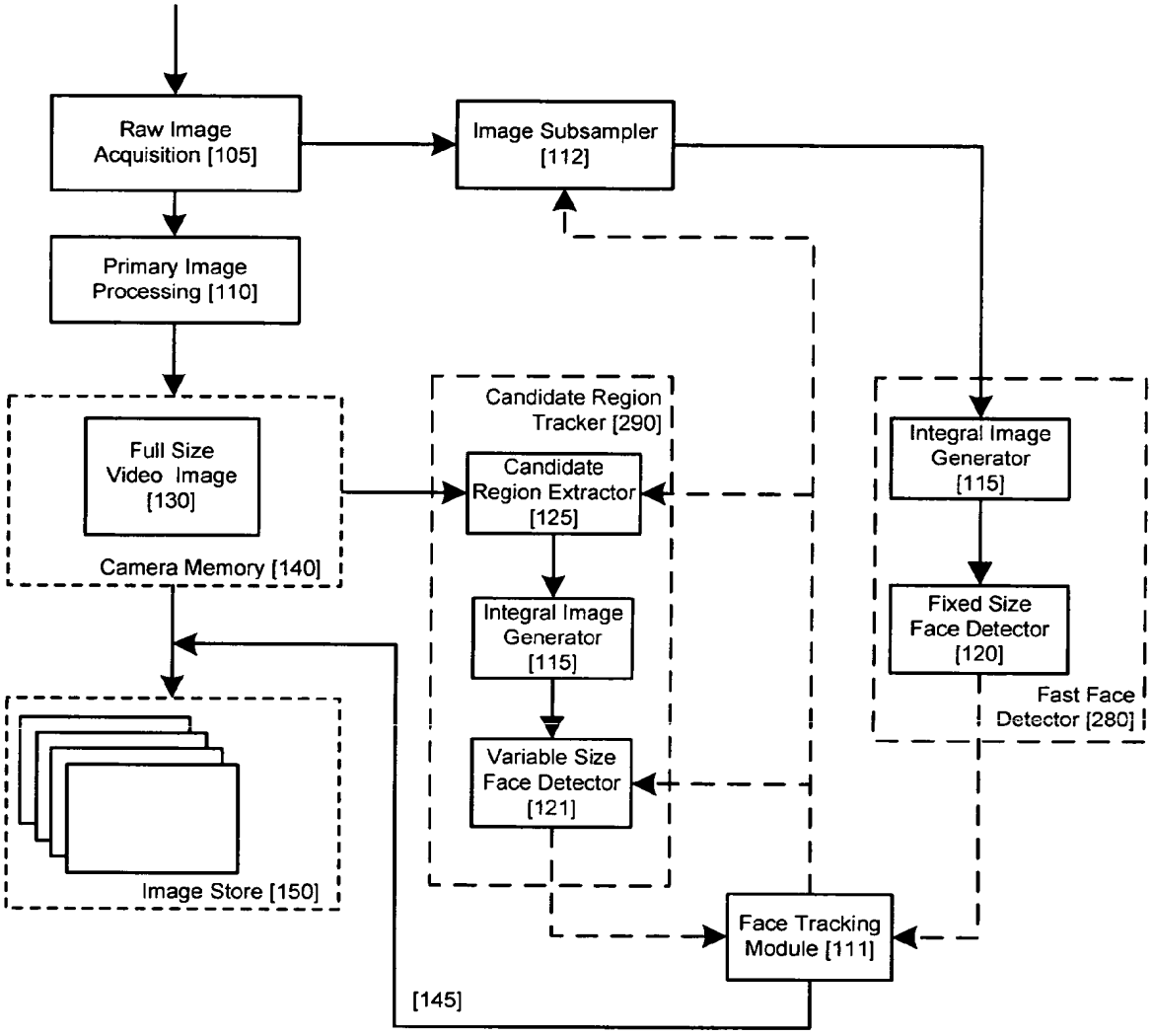

Raw Image Acquisition [105]
Image Subsampler [112]
Primary Image Processing [110]
Full Size Video Image [130]
Camera Memory [140]
Candidate Region Tracker [290]
Candidate Region Extractor [125]
Integral Image Generator [115]
Variable Size Face Detector [121]
Integral Image Generator [115]
Fixed Size Face Detector [120]
Fast Face Detector [280]
Image Store [150]
Face Tracking Module [111]
[145]

Figure 1

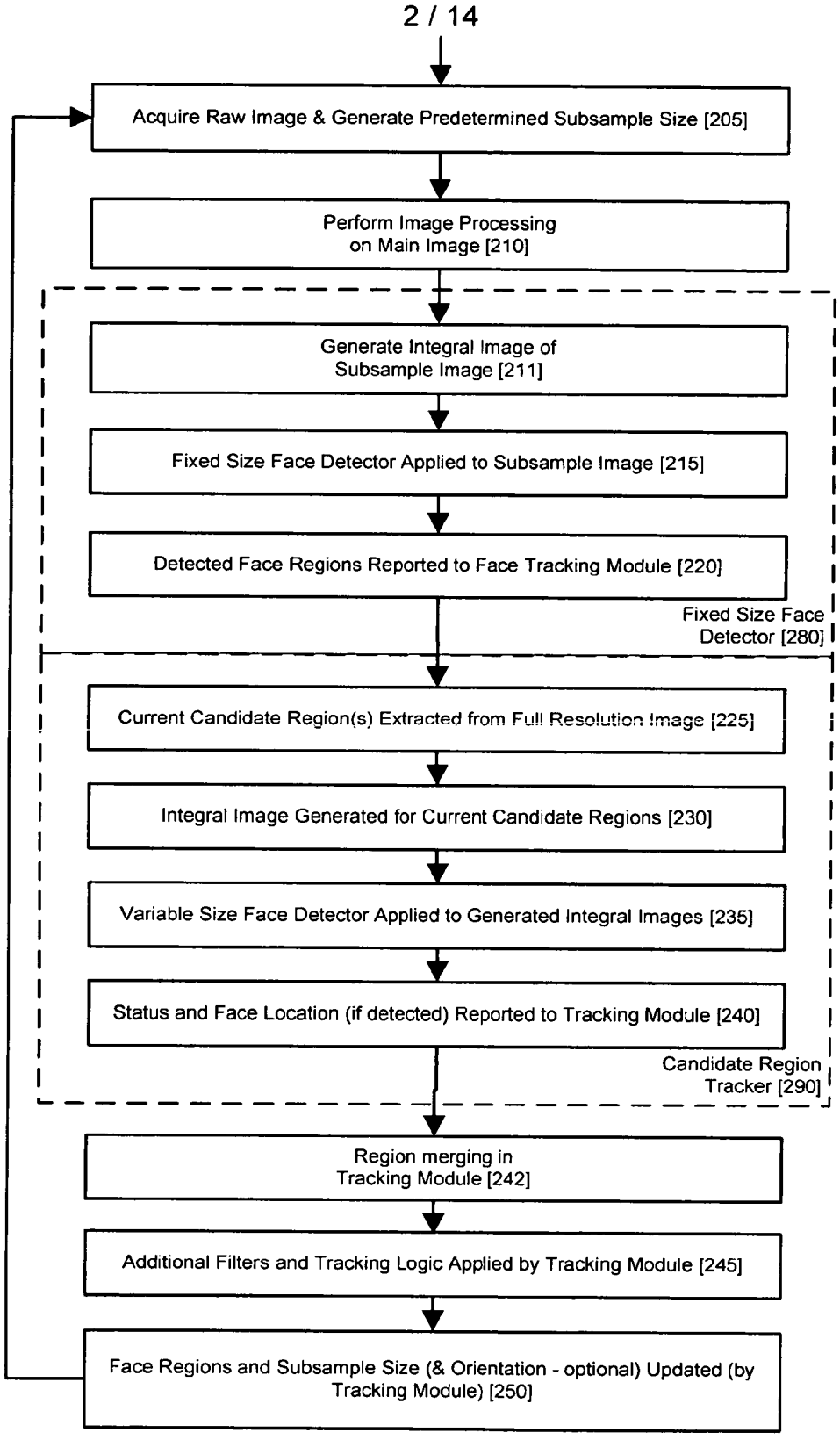

Acquire Raw Image & Generate Predetermined Subsample Size [205]
Perform Image Processing on Main Image [210]
Generate Integral Image of Subsample Image [211]
Fixed Size Face Detector Applied to Subsample Image [215]
Detected Face Regions Reported to Face Tracking Module [220]
Fixed Size Face Detector [280]
Current Candidate Region(s) Extracted from Full Resolution Image [225]
Integral Image Generated for Current Candidate Regions [230]
Variable Size Face Detector Applied to Generated Integral Images [235]
Status and Face Location (if detected) Reported to Tracking Module [240]
Candidate Region Tracker [290]
Region merging in Tracking Module [242]
Additional Filters and Tracking Logic Applied by Tracking Module [245]
Face Regions and Subsample Size (& Orientation - optional) Updated (by Tracking Module) [250]

Figure 2

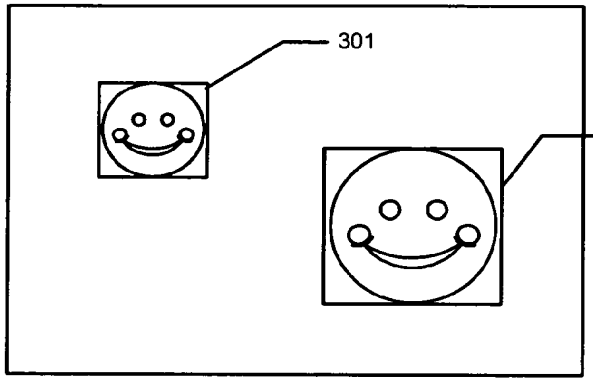


Figure 3A

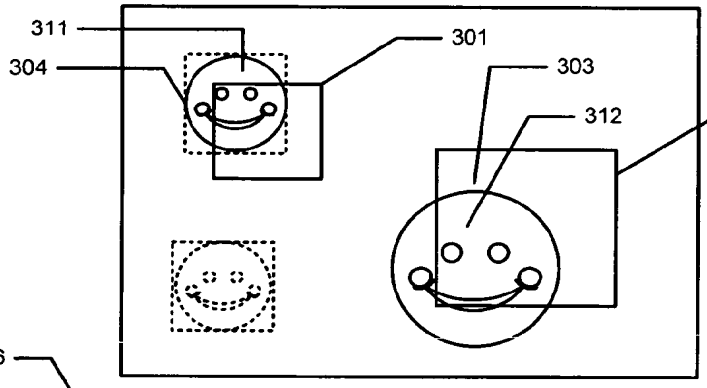


Figure 3B

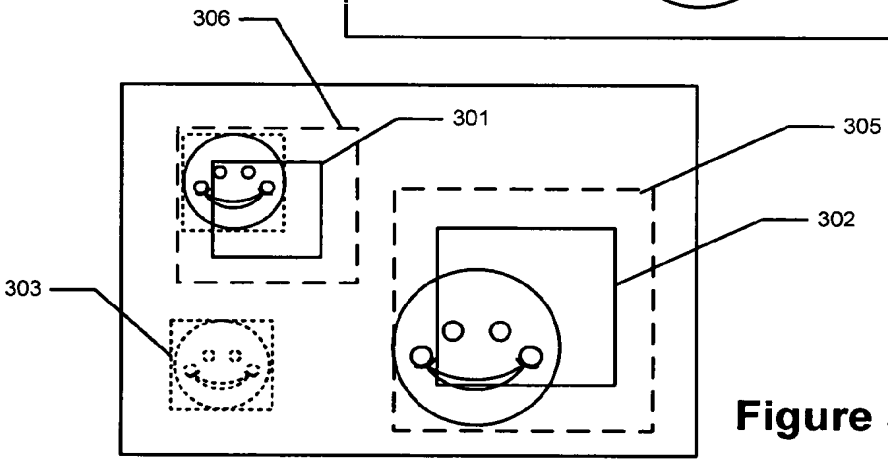


Figure 3C

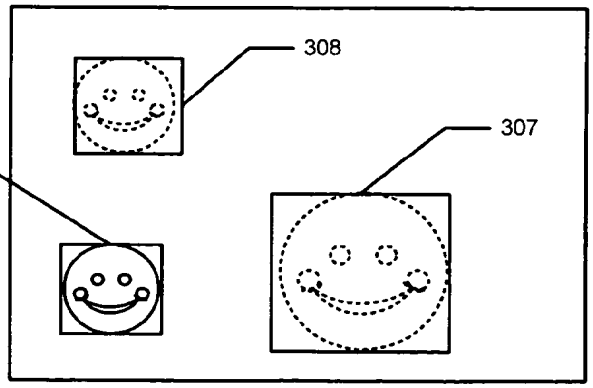


Figure 3D

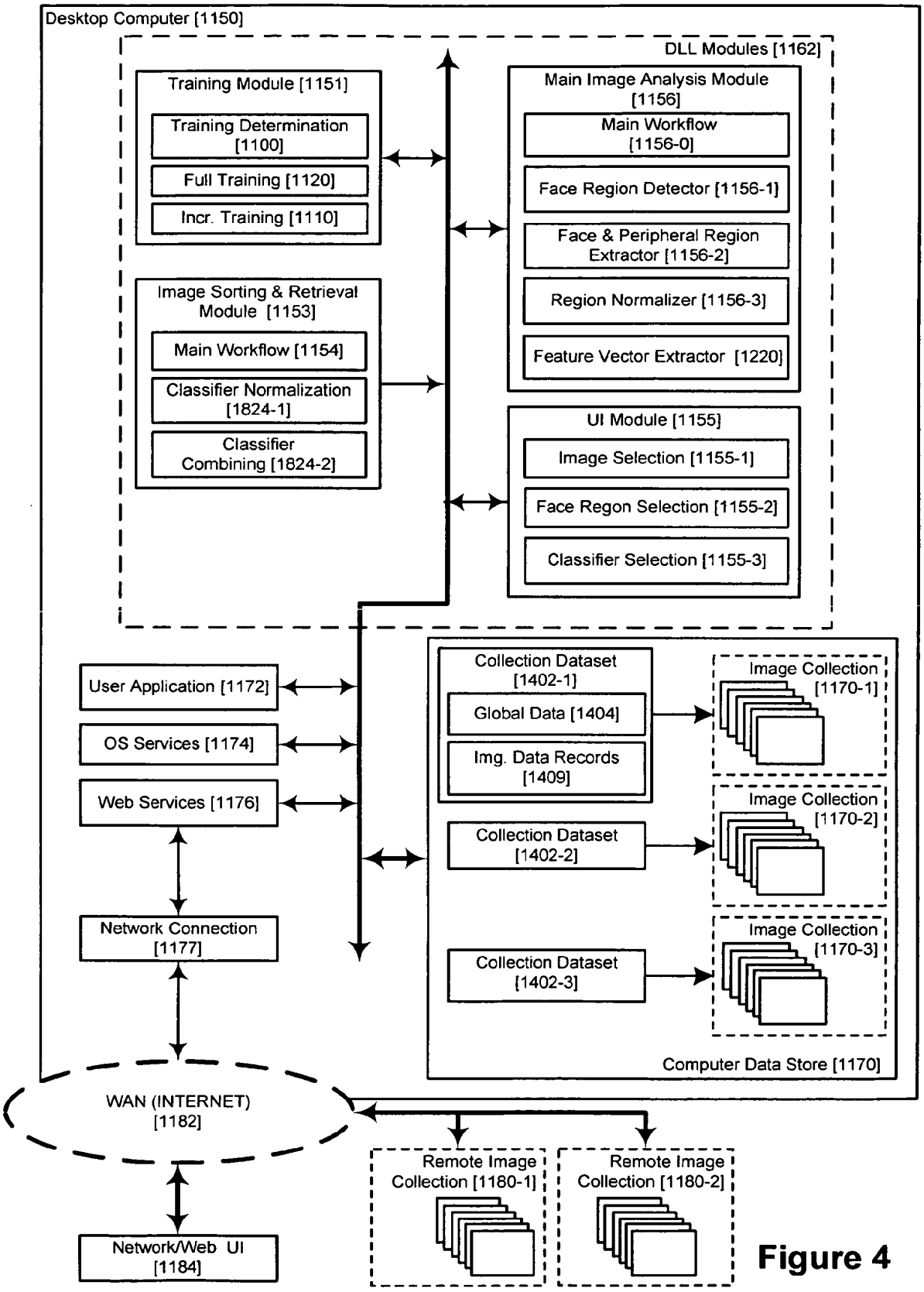

Desktop Computer [1150]
DLL Modules [1162]
Training Module [1151]
Training Determination [1100]
Full Training [1120]
Incr. Training [1110]
Main Image Analysis Module [1156]
Main Workflow [1156-0]
Face Region Detector [1156-1]
Face & Peripheral Region Extractor [1156-2]
Region Normalizer [1156-3]
Feature Vector Extractor [1220]
Image Sorting & Retrieval Module [1153]
Main Workflow [1154]
Classifier Normalization [1824-1]
Classifier Combining [1824-2]
UI Module [1155]
Image Selection [1155-1]
Face Regon Selection [1155-2]
Classifier Selection [1155-3]
User Application [1172]
OS Services [1174]
Web Services [1176]
Network Connection [1177]
Collection Dataset [1402-1]
Global Data [1404]
Img. Data Records [1409]
Collection Dataset [1402-2]
Collection Dataset [1402-3]
Image Collection [1170-1]
Image Collection [1170-2]
Image Collection [1170-3]
Computer Data Store [1170]
WAN (INTERNET) [1182]
Remote Image Collection [1180-1]
Remote Image Collection [1180-2]
Network/Web UI [1184]

Figure 4

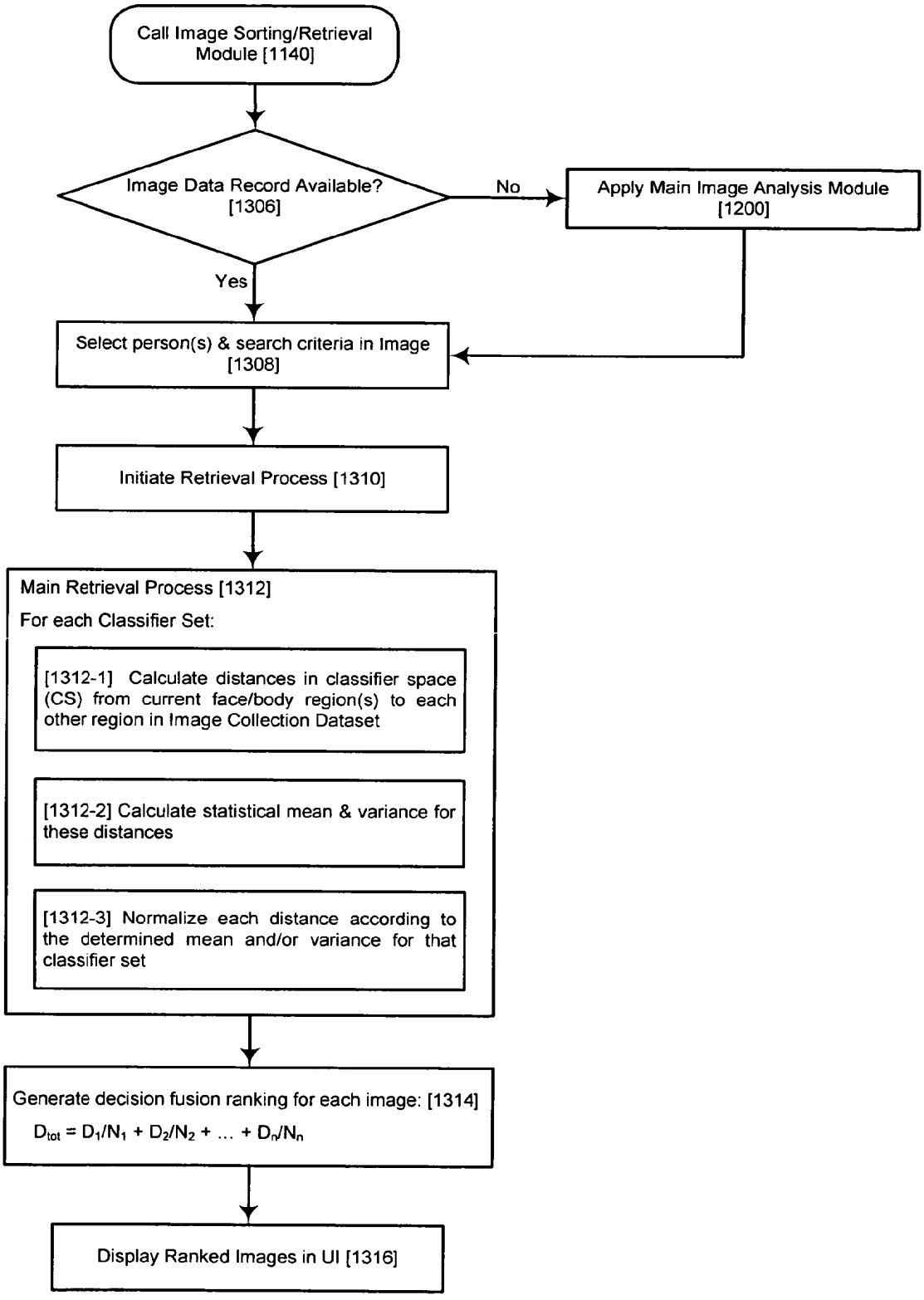

Call Image Sorting/Retrieval Module [1140]
Image Data Record Available? [1306]
No
Apply Main Image Analysis Module [1200]
Yes
Select person(s) & search criteria in Image [1308]
Initiate Retrieval Process [1310]
Main Retrieval Process [1312]
For each Classifier Set:
[1312-1] Calculate distances in classifier space (CS) from current face/body region(s) to each other region in Image Collection Dataset
[1312-2] Calculate statistical mean & variance for these distances
[1312-3] Normalize each distance according to the determined mean and/or variance for that classifier set
Generate decision fusion ranking for each image: [1314]
$D_{tot} = D_1/N_1 + D_2/N_2 + \ldots + D_n/N_n$
Display Ranked Images in UI [1316]

Figure 5

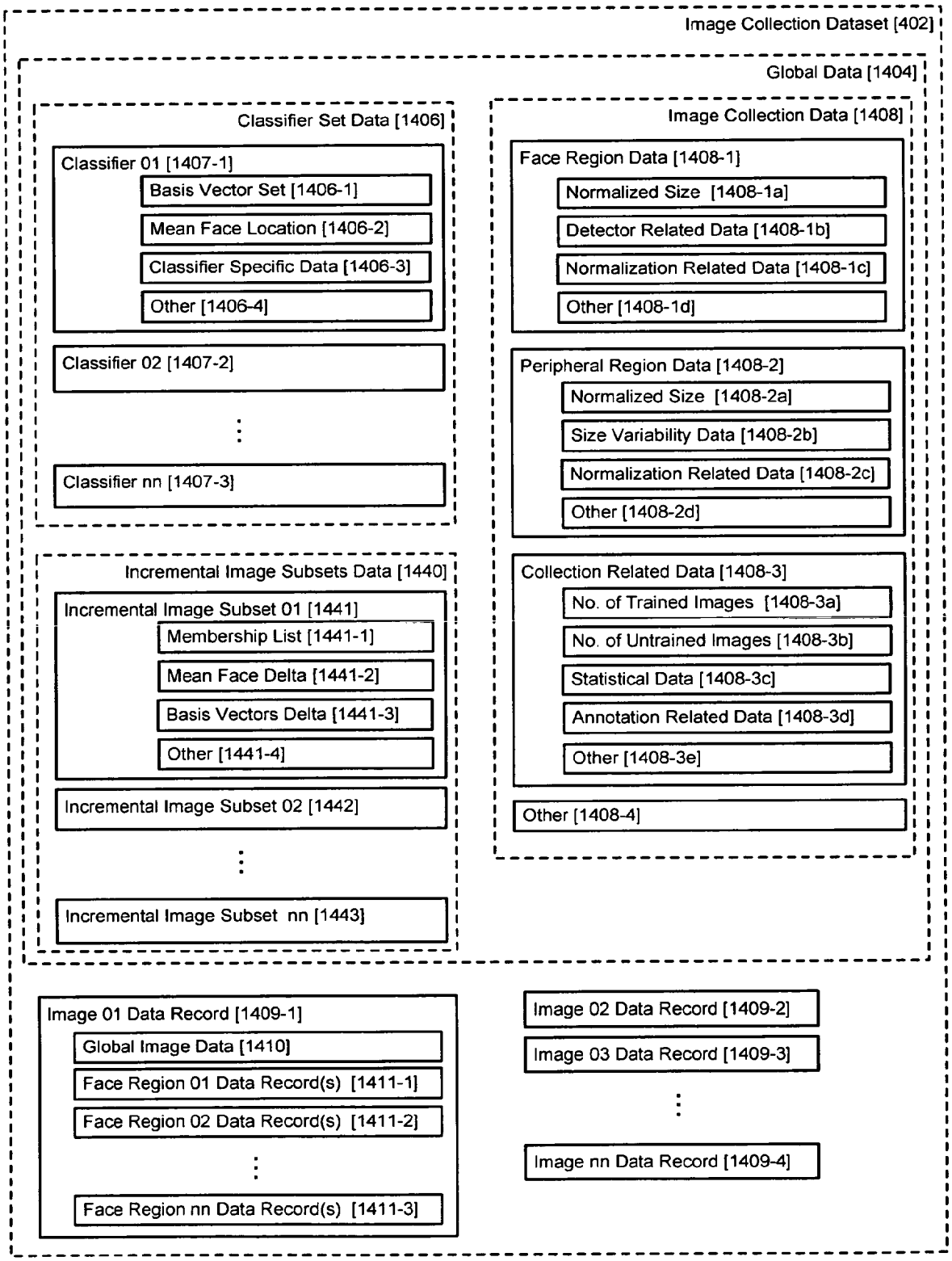


Image Collection Dataset [402]
Global Data [1404]
Classifier Set Data [1406]
Classifier 01 [1407-1]
Basis Vector Set [1406-1]
Mean Face Location [1406-2]
Classifier Specific Data [1406-3]
Other [1406-4]
Classifier 02 [1407-2]
Classifier nn [1407-3]
Incremental Image Subsets Data [1440]
Incremental Image Subset 01 [1441]
Membership List [1441-1]
Mean Face Delta [1441-2]
Basis Vectors Delta [1441-3]
Other [1441-4]
Incremental Image Subset 02 [1442]
Incremental Image Subset nn [1443]
Image Collection Data [1408]
Face Region Data [1408-1]
Normalized Size [1408-1a]
Detector Related Data [1408-1b]
Normalization Related Data [1408-1c]
Other [1408-1d]
Peripheral Region Data [1408-2]
Normalized Size [1408-2a]
Size Variability Data [1408-2b]
Normalization Related Data [1408-2c]
Other [1408-2d]
Collection Related Data [1408-3]
No. of Trained Images [1408-3a]
No. of Untrained Images [1408-3b]
Statistical Data [1408-3c]
Annotation Related Data [1408-3d]
Other [1408-3e]
Other [1408-4]
Image 01 Data Record [1409-1]
Global Image Data [1410]
Face Region 01 Data Record(s) [1411-1]
Face Region 02 Data Record(s) [1411-2]
Face Region nn Data Record(s) [1411-3]
Image 02 Data Record [1409-2]
Image 03 Data Record [1409-3]
Image nn Data Record [1409-4]

Figure 6A

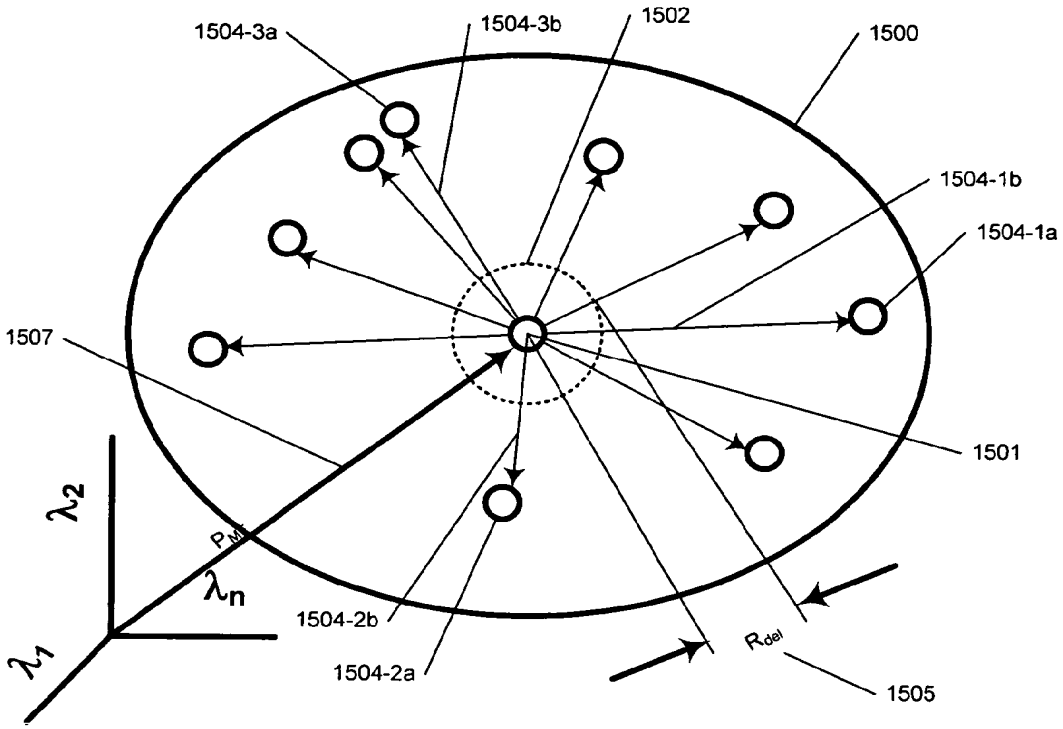


Figure 6B

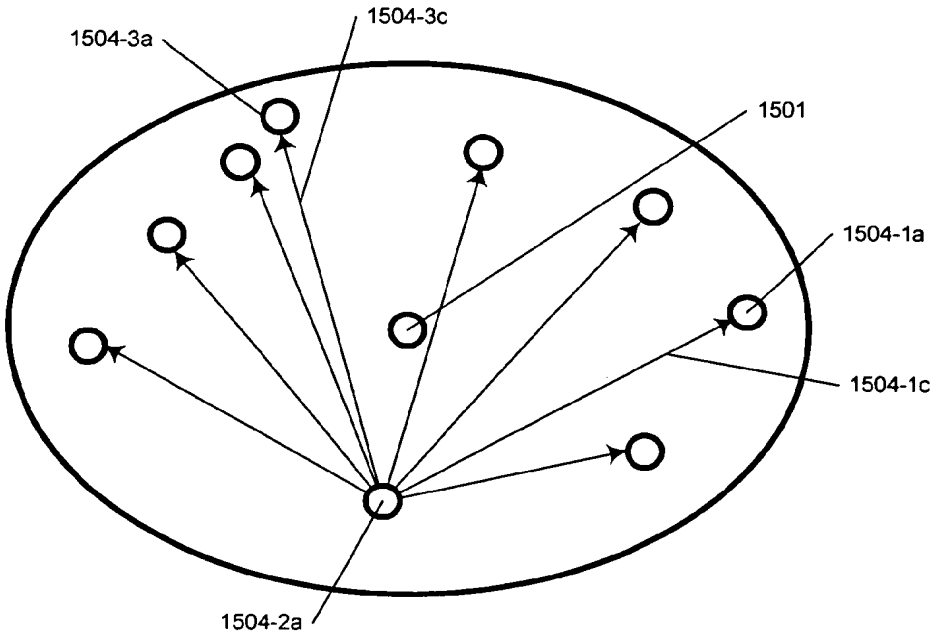


Figure 6C

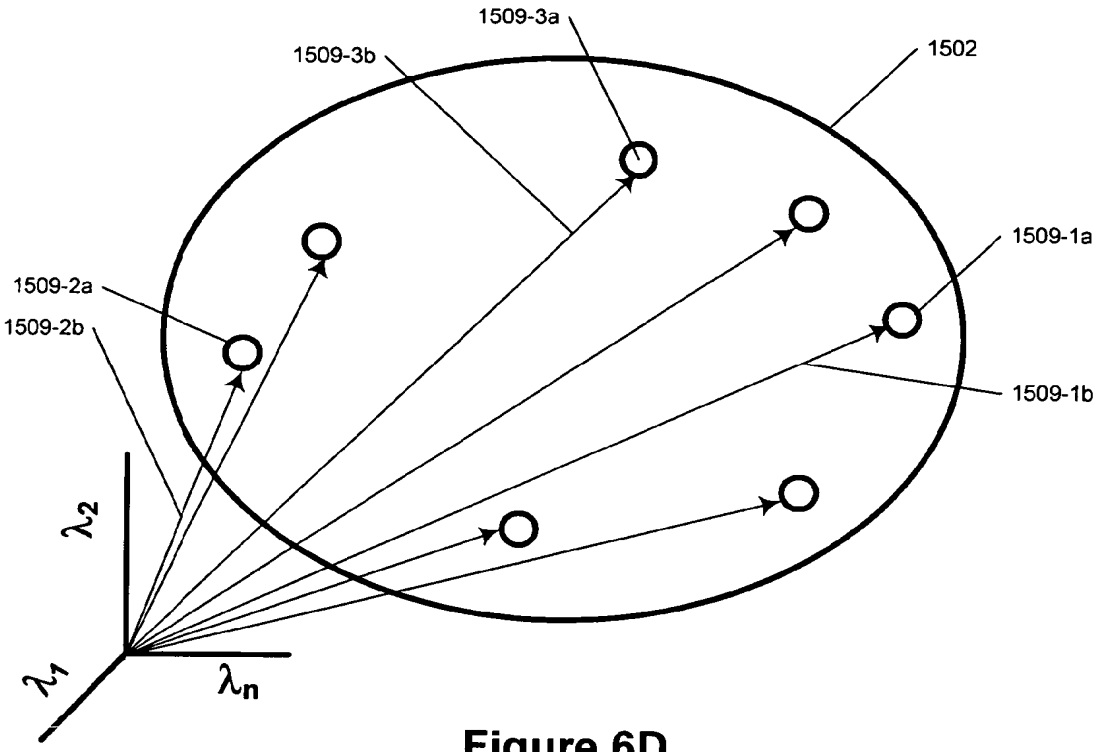


Figure 6D

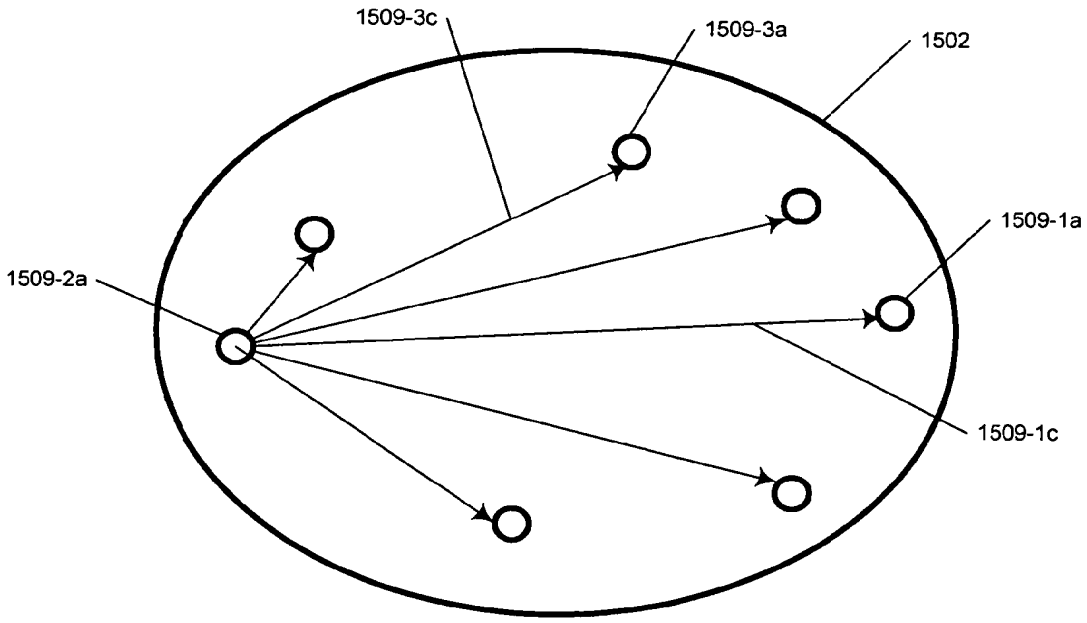
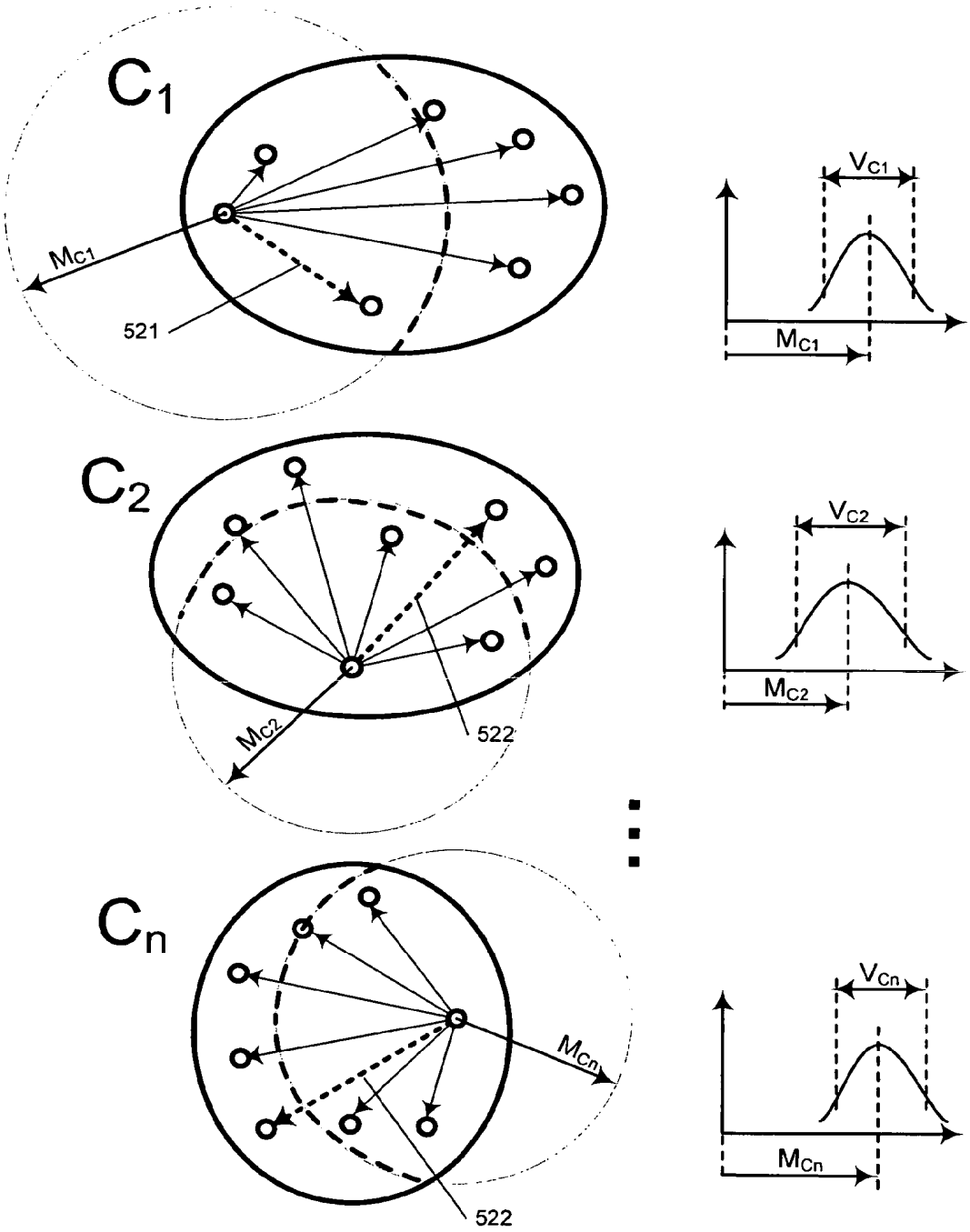


Figure 6E



$$D_{tot} = D_1/M_{C1} + D_2/M_{C2} + \ldots + D_n/M_{Cn}$$
$$= D'_1 + D'_2 + \ldots + D'_n$$

Figure 6F

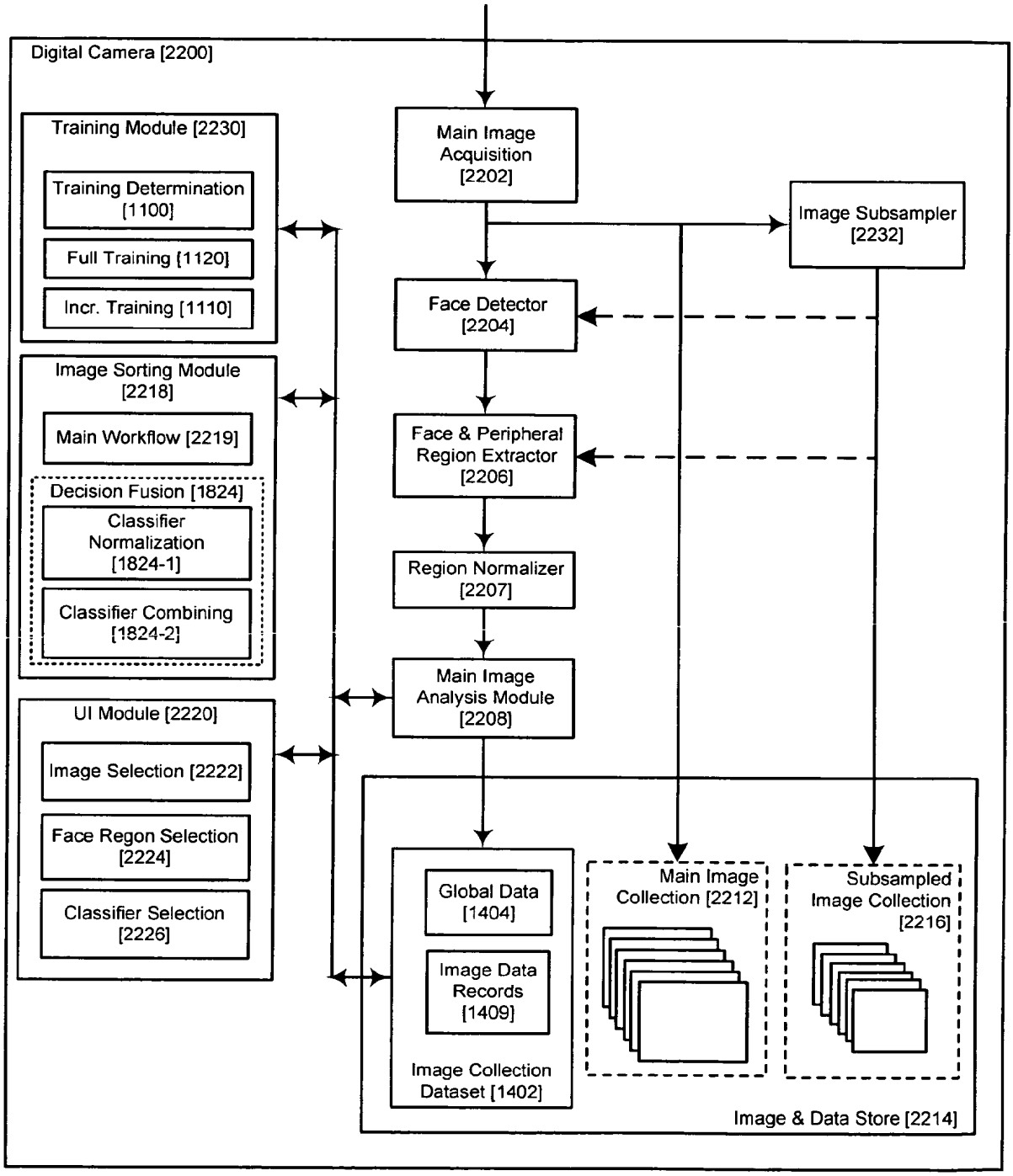

Digital Camera [2200]
Training Module [2230]
Training Determination [1100]
Full Training [1120]
Incr. Training [1110]
Image Sorting Module [2218]
Main Workflow [2219]
Decision Fusion [1824]
Classifier Normalization [1824-1]
Classifier Combining [1824-2]
UI Module [2220]
Image Selection [2222]
Face Regon Selection [2224]
Classifier Selection [2226]
Main Image Acquisition [2202]
Image Subsampler [2232]
Face Detector [2204]
Face & Peripheral Region Extractor [2206]
Region Normalizer [2207]
Main Image Analysis Module [2208]
Global Data [1404]
Image Data Records [1409]
Image Collection Dataset [1402]
Main Image Collection [2212]
Subsampled Image Collection [2216]
Image & Data Store [2214]

Figure 7

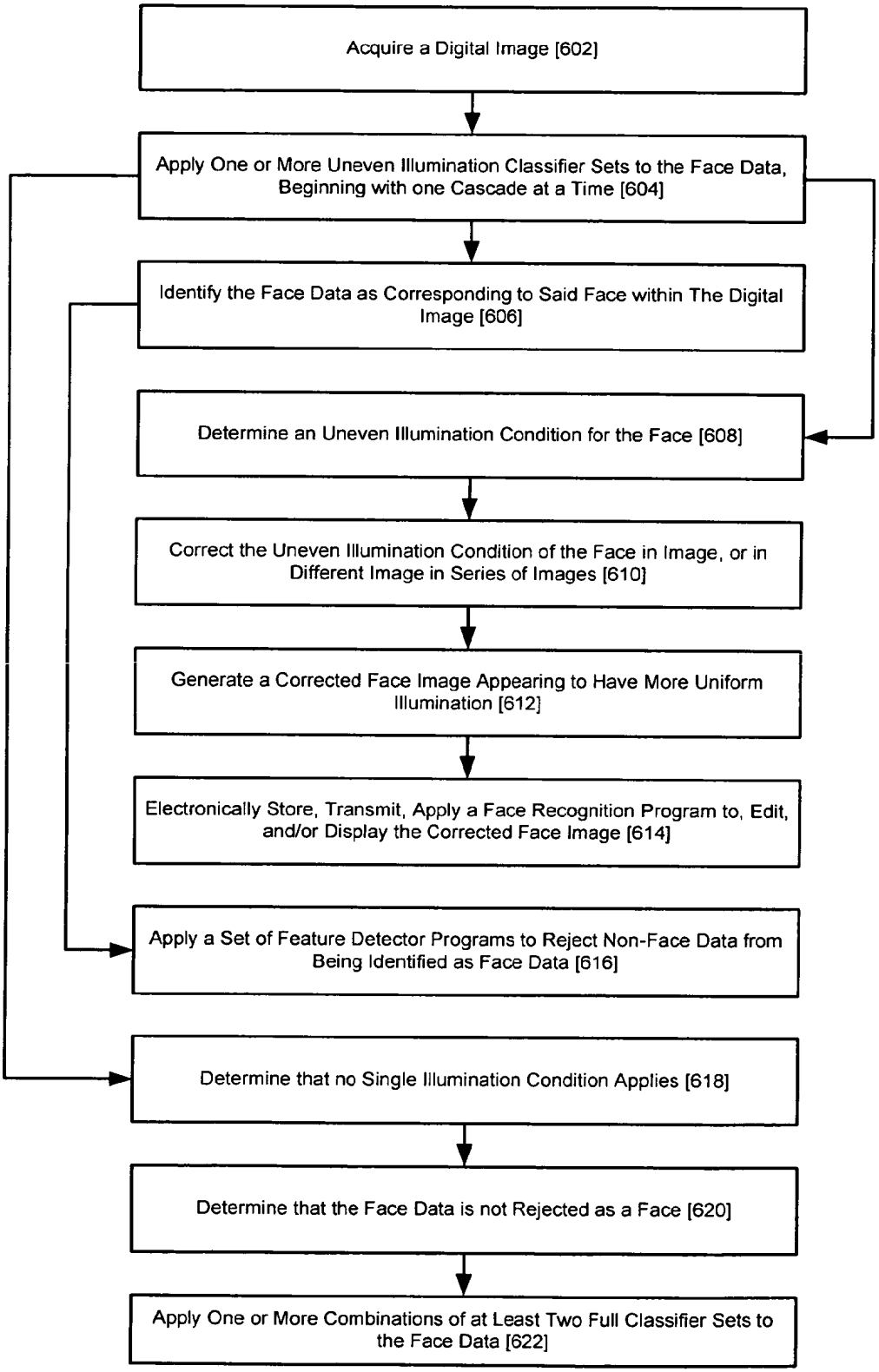

Acquire a Digital Image [602]
Apply One or More Uneven Illumination Classifier Sets to the Face Data, Beginning with one Cascade at a Time [604]
Identify the Face Data as Corresponding to Said Face within The Digital Image [606]
Determine an Uneven Illumination Condition for the Face [608]
Correct the Uneven Illumination Condition of the Face in Image, or in Different Image in Series of Images [610]
Generate a Corrected Face Image Appearing to Have More Uniform Illumination [612]
Electronically Store, Transmit, Apply a Face Recognition Program to, Edit, and/or Display the Corrected Face Image [614]
Apply a Set of Feature Detector Programs to Reject Non-Face Data from Being Identified as Face Data [616]
Determine that no Single Illumination Condition Applies [618]
Determine that the Face Data is not Rejected as a Face [620]
Apply One or More Combinations of at Least Two Full Classifier Sets to the Face Data [622]

Figure 8

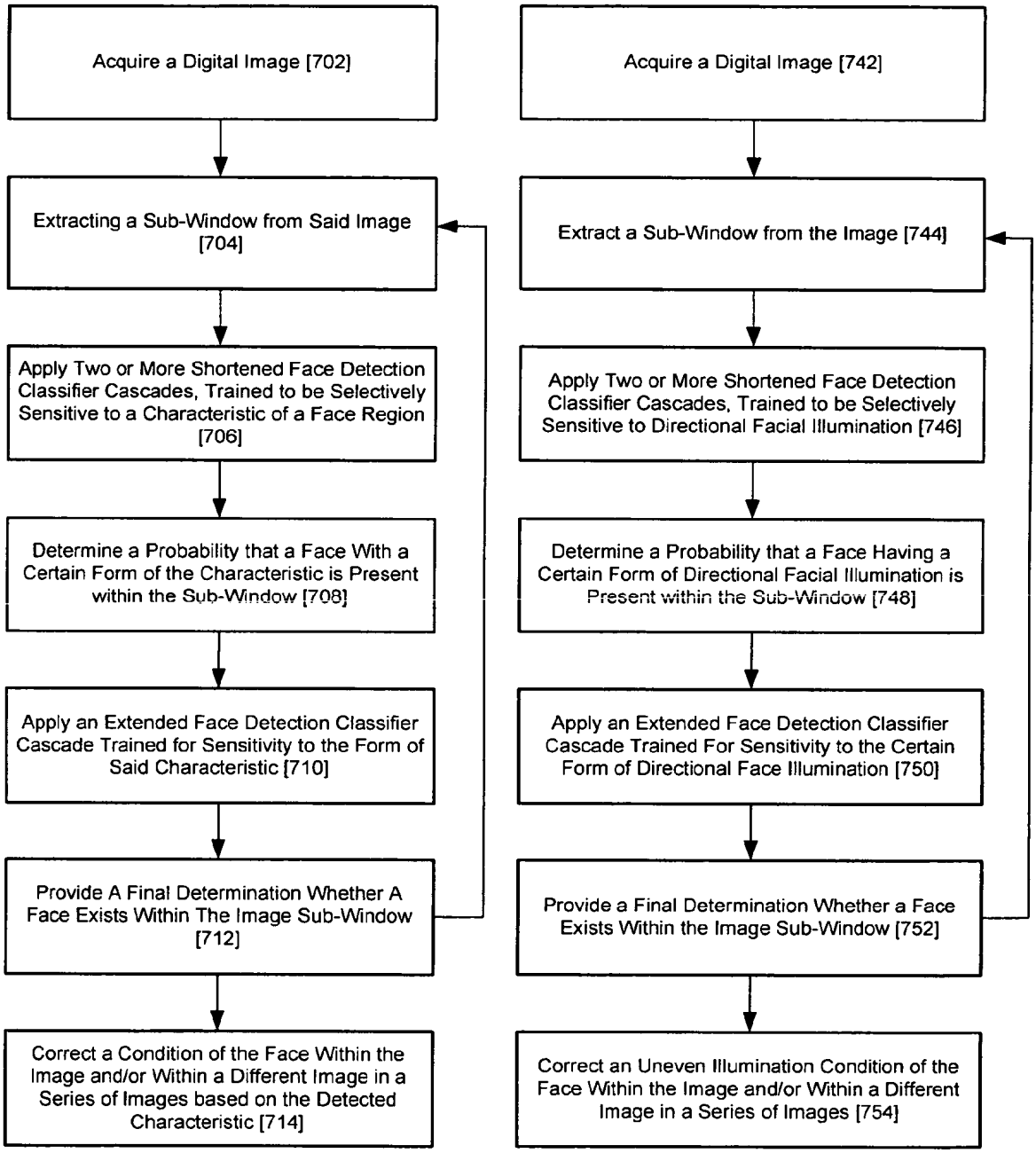


Figure 9A **Figure 9B**

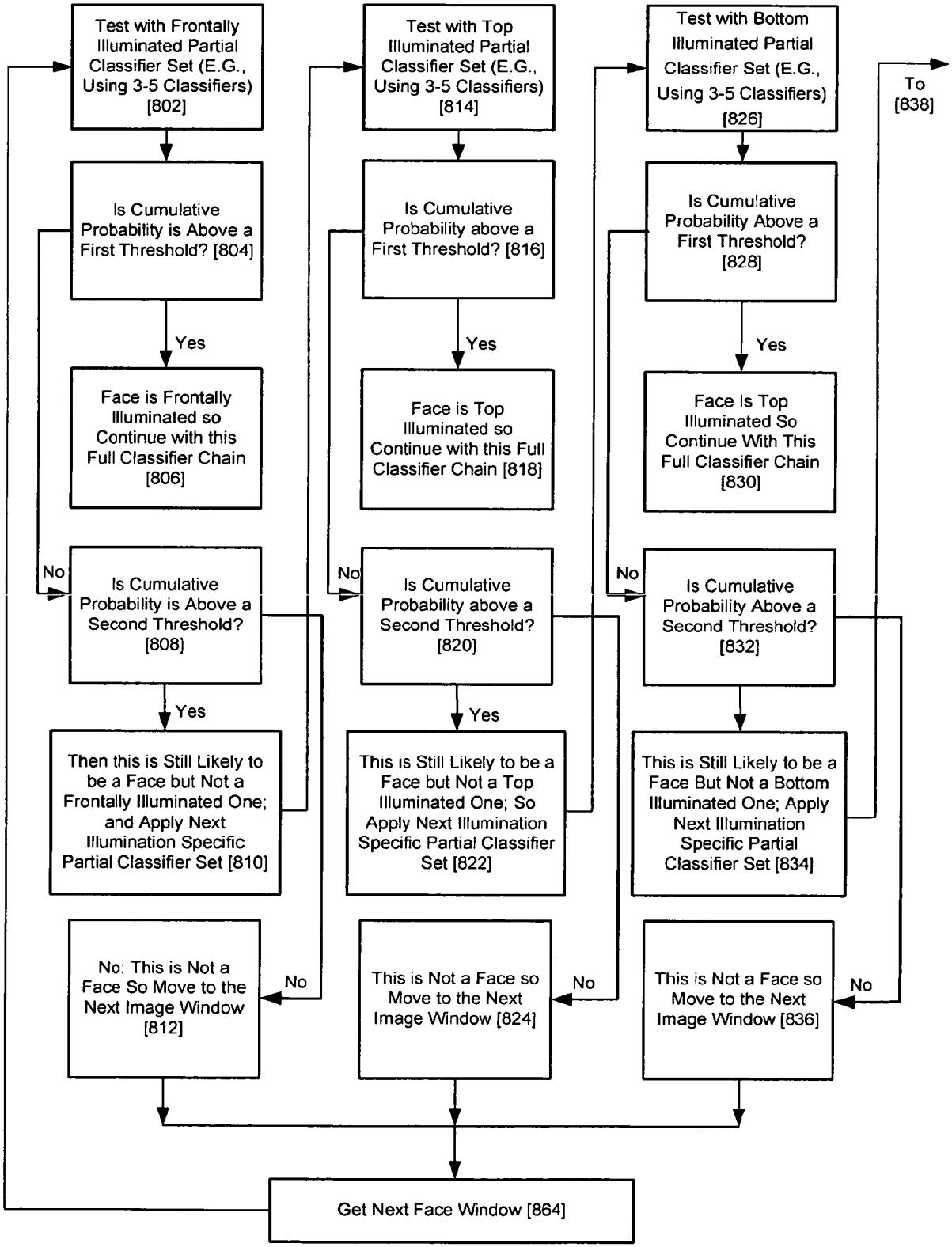

Test with Frontally Illuminated Partial Classifier Set (E.G., Using 3-5 Classifiers) [802]
Is Cumulative Probability is Above a First Threshold? [804]
Yes
Face is Frontally Illuminated so Continue with this Full Classifier Chain [806]
No
Is Cumulative Probability is Above a Second Threshold? [808]
Yes
Then this is Still Likely to be a Face but Not a Frontally Illuminated One; and Apply Next Illumination Specific Partial Classifier Set [810]
No: This is Not a Face So Move to the Next Image Window [812]
No
Test with Top Illuminated Partial Classifier Set (E.G., Using 3-5 Classifiers) [814]
Is Cumulative Probability above a First Threshold? [816]
Yes
Face is Top Illuminated so Continue with this Full Classifier Chain [818]
No
Is Cumulative Probability above a Second Threshold? [820]
Yes
This is Still Likely to be a Face but Not a Top Illuminated One; So Apply Next Illumination Specific Partial Classifier Set [822]
This is Not a Face so Move to the Next Image Window [824]
No
Test with Bottom Illuminated Partial Classifier Set (E.G., Using 3-5 Classifiers) [826]
Is Cumulative Probability Above a First Threshold? [828]
Yes
Face Is Top Illuminated So Continue With This Full Classifier Chain [830]
No
Is Cumulative Probability Above a Second Threshold? [832]
Yes
This is Still Likely to be a Face But Not a Bottom Illuminated One; Apply Next Illumination Specific Partial Classifier Set [834]
This is Not a Face so Move to the Next Image Window [836]
No
To [838]
Get Next Face Window [864]

Figure 10A

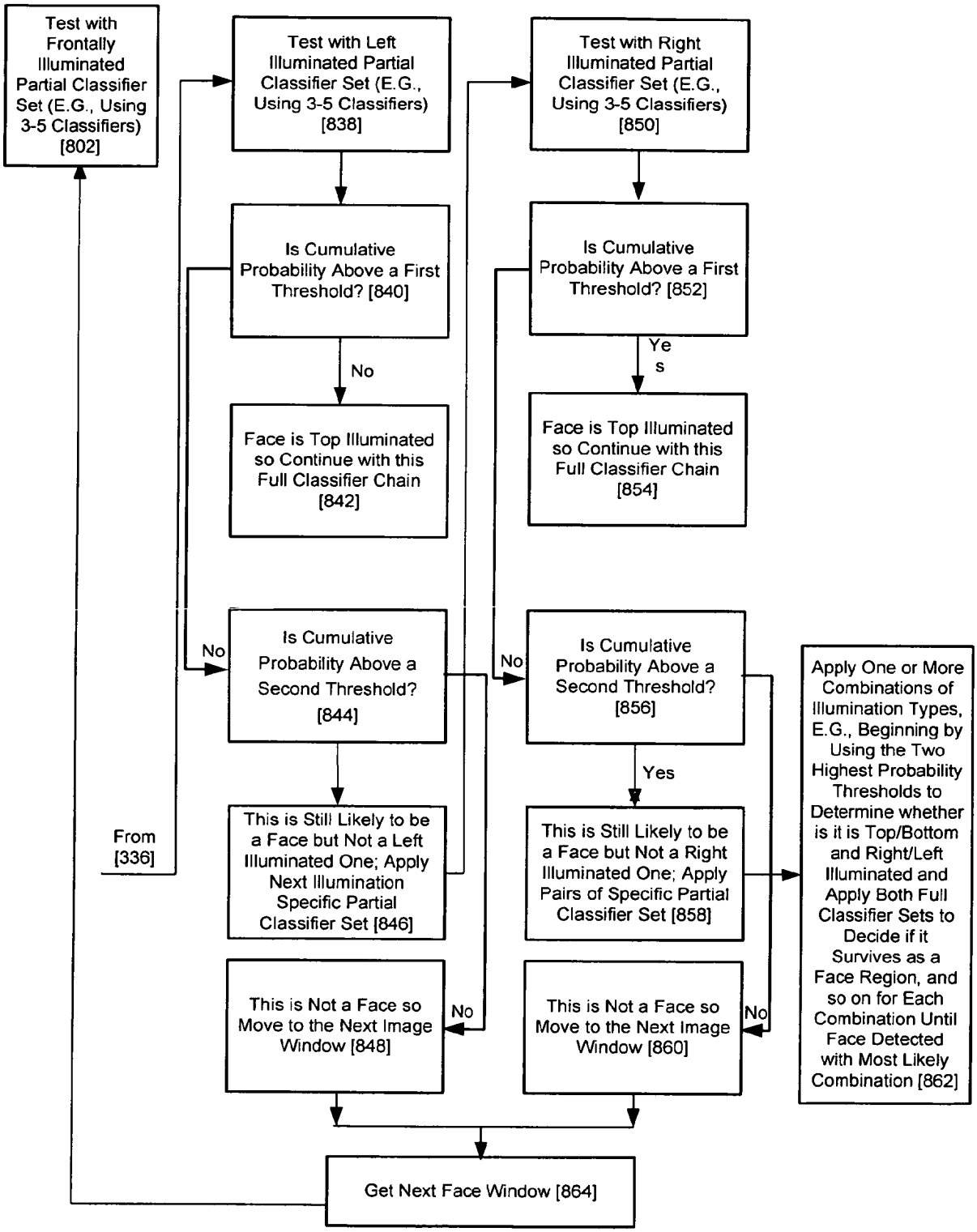

Test with Frontally Illuminated Partial Classifier Set (E.G., Using 3-5 Classifiers) [802]
Test with Left Illuminated Partial Classifier Set (E.G., Using 3-5 Classifiers) [838]
Test with Right Illuminated Partial Classifier Set (E.G., Using 3-5 Classifiers) [850]
Is Cumulative Probability Above a First Threshold? [840]
Is Cumulative Probability Above a First Threshold? [852]
No
Ye s
Face is Top Illuminated so Continue with this Full Classifier Chain [842]
Face is Top Illuminated so Continue with this Full Classifier Chain [854]
No
Is Cumulative Probability Above a Second Threshold? [844]
No
Is Cumulative Probability Above a Second Threshold? [856]
Yes
From [336]
This is Still Likely to be a Face but Not a Left Illuminated One; Apply Next Illumination Specific Partial Classifier Set [846]
This is Still Likely to be a Face but Not a Right Illuminated One; Apply Pairs of Specific Partial Classifier Set [858]
Apply One or More Combinations of Illumination Types, E.G., Beginning by Using the Two Highest Probability Thresholds to Determine whether is it is Top/Bottom and Right/Left Illuminated and Apply Both Full Classifier Sets to Decide if it Survives as a Face Region, and so on for Each Combination Until Face Detected with Most Likely Combination [862]
This is Not a Face so Move to the Next Image Window [848]
No
This is Not a Face so Move to the Next Image Window [860]
No
Get Next Face Window [864]

Figure 10B